



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICAS PÚBLICAS, ESTRATÉGIAS E
DESENVOLVIMENTO

MIGUEL BASTOS ALVARENGA

**O DEBATE REGULATÓRIO SOBRE MINERAÇÃO DE TEXTOS E DADOS E DI-
REITO AUTORAL NA PESQUISA BRASILEIRA**

Rio de Janeiro

2025

MIGUEL BASTOS ALVARENGA

O DEBATE REGULATÓRIO SOBRE MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL NA PESQUISA BRASILEIRA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Políticas Públicas, Estratégias e Desenvolvimento, Instituto de Economia, Universidade Federal do Rio de Janeiro – UFRJ, na área de Inovação, Propriedade Intelectual e Desenvolvimento, como requisito parcial à obtenção do título de Doutor em Políticas Públicas, Estratégias e Desenvolvimento.

Orientador: Dr. Allan Rocha de Souza

Rio de Janeiro

2025

FICHA CATALOGRÁFICA

A473d Alvarenga, Miguel Bastos.
O debate regulatório sobre mineração de textos e dados e direito autoral na
pesquisa brasileira / Miguel Bastos Alvarenga . - 2025.
296 f.

Orientador: Allan Rocha de Souza.
Tese (doutorado) – Universidade Federal do Rio de Janeiro, Instituto de
Economia, Programa de Pós-Graduação em Políticas Públicas, Estratégias e
Desenvolvimento, 2025.
Bibliografia: f. 249 – 289.

1. Mineração de dados. 2. Direito autoral. 3. Pesquisa. 4. Inteligência artificial.
I. Souza, Allan Rocha de, orient. II. Universidade Federal do Rio de Janeiro.
Instituto de Economia. III. Título.

CDD 005.7

MIGUEL BASTOS ALVARENGA

O DEBATE REGULATÓRIO SOBRE MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL NA PESQUISA BRASILEIRA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Políticas Públicas, Estratégias e Desenvolvimento, Instituto de Economia, Universidade Federal do Rio de Janeiro – UFRJ, na área de Inovação, Propriedade Intelectual e Desenvolvimento, como requisito parcial à obtenção do título de Doutor em Políticas Públicas, Estratégias e Desenvolvimento.

Orientador: Dr. Allan Rocha de Souza

Aprovado em: 29 de abril de 2025

Dr. Allan Rocha de Souza – PPED/UFRJ

Dra. Maria Tereza Leopardi Mello – PPED/UFRJ

Dra. Mariana Valente – University of St. Gallen

Dr. Luca Schirru – PUC-Rio

Dr. Nicolo Zingales – Fundação Getúlio Vargas

AGRADECIMENTOS

Ao longo desses anos de estudo e trabalho – na academia e no Direito Autoral –, aprendi que produzir conhecimento é lidar, ao mesmo tempo, com dois lados de uma mesma moeda. O primeiro é o desejo de ter um trabalho reconhecido como seu; é o orgulho (e a obrigação) de ter seu nome destacado dos demais pela contribuição que você, e só você, trouxe. O segundo é a constatação de que, paradoxalmente, seu trabalho não é só seu; é fruto do seu diálogo com aqueles que vieram antes e os que vêm junto de você. É também produto daqueles que lhe apoiaram ao longo da vida e da carreira. É a essas pessoas que dedico esses agradecimentos.

Ao meu orientador, Allan Rocha de Souza, com quem tenho e sempre tive muito orgulho de trabalhar. Obrigado pelas lições, pelos conselhos, pela camaradagem e, principalmente, por ver potencial em mim e me encorajar a sempre ir mais longe. É um privilégio estar sob seus cuidados.

A Mariana Valente, que supervisionou meu trabalho durante o doutorado-sanduíche no exterior. É normal nos desaconselharem a conhecer nossos heróis, mas, neste caso, eles não poderiam estar mais errados. Obrigado pela acolhida, pelo apoio e por ter não só ajudado a realizar um sonho profissional meu, mas também a dar algumas guinadas nesta tese.

Aos membros da banca examinadora, pelos apontamentos, críticas e sugestões que, sem sombra de dúvidas, contribuíram enormemente para o aprimoramento deste trabalho.

Aos colegas e amigos do Instituto Brasileiro de Direitos Autorais, pela amizade, pelas trocas e por todo o apoio ao longo deste trabalho. Seja compartilhando material, seja escutando algumas ideias malucas e me acalmando nas crises de ansiedade – que, como vocês sabem, não foram poucas!

A Sean Flynn, Mike Palmedo e a toda a equipe do *Program of Information Justice and Intellectual Property* da *American University*, pelos diálogos, pelas trocas, pela amizade e por todo o apoio intelectual e pessoal nesses últimos anos.

A todo o corpo docente, discente e técnico da Universidade de St. Gallen, pela acolhida, pela troca de experiências e por todo o apoio durante meu período de estudos no doutorado-sanduíche. Espero poder revê-los quando o clima estiver menos frio!

“Aos mestres, com carinho”: todos os professores e colegas de academia, governo e sociedade civil com quem aprendi tudo o que sei e com muitos dos quais tive o privilégio de pessoalmente conversar, compartilhar conhecimento e até comida! Este trabalho é, em parte, uma homenagem minha a todos vocês.

À CAPES, pelo apoio financeiro e de infraestrutura ao longo do doutorado, tanto no Brasil como durante o doutorado sanduíche no exterior.

Ao Laboratório de Humanidades Digitais da PUC-Rio, pelo apoio de infraestrutura na pesquisa de campo.

Aos meus pais, que sempre garantiram que nada me faltasse. Comida, amor, apoio... Eu não teria sido capaz de fazer nada disto aqui sem vocês. Mãe, saiba que, mesmo à distância, a senhora nunca estará sozinha; eu te amo e sou grato por tudo. Espero que a senhora fique orgulhosa de mim! Pai, onde você estiver, quero que saiba que carrego muito do senhor comigo até nas coisas pequenas. Desculpa não ter dado tempo de o senhor me ver aqui agora, mas espero que a vista daí de cima seja boa!

À Luísa, meu grande amor de mais de onze anos e minha melhor amiga de ainda mais tempo. Obrigado por me fazer gostar da rotina, de comédia romântica, de cores fortes... Por ter curado meu medo de cachorro. Por ter me introduzido aos estudos de ciência e tecnologia. Por sempre me apoiar – e puxar minha orelha quando preciso. Por me fazer uma pessoa melhor. Por ser simplesmente você.

À minha família: minha primeira casa, meus primeiros amigos e aqueles que, nos momentos alegres e difíceis, mostraram que nunca estive sozinho e que sempre tive um porto seguro para me apoiar, para brincar e para me reconectar comigo mesmo. E a meus sogros e toda a sua família, por todo o apoio, carinho e brincadeiras – por me fazerem sentir também parte de vocês.

A todos os meus colegas e amigos em Belém, no Rio, no restante do Brasil e no exterior, presenciais e online – que estiveram comigo nos dias mais felizes e nos mais tristes, me acolheram quando estava perdido e sozinho e que, de uma forma ou outra, foram grandes responsáveis pela minha trajetória não apenas profissional, mas também pessoal. Sem vocês, cada um pelas mais variadas razões, eu não estaria aqui hoje, e sou muito feliz por compartilhar este mundo com vocês.

E a todos que, direta ou indiretamente, participaram deste trabalho.

Database, database

I'm struggling in the database, whoa-oh

It doesn't even matter if there is no hope

As the madness of the system grows

(MAN WITH A MISSION feat. Takuma, 2013)

RESUMO

Alvarenga, M. B. **O DEBATE REGULATÓRIO SOBRE MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL NA PESQUISA BRASILEIRA**. Orientador: Allan Rocha de Souza. 2025. Tese (Doutorado em Políticas Públicas, Estratégias e Desenvolvimento) – Instituto de Economia – IE, Universidade Federal do Rio de Janeiro, 2025.

A presente tese de doutorado tem por objetivo avaliar se o atual debate e proposta regulatórios sobre mineração de textos e dados atendem às necessidades do ecossistema de pesquisa, considerando as políticas públicas de inovação e o ordenamento jurídico brasileiro. Tal pesquisa é motivada pelo crescimento no uso de ferramentas automatizadas de coleta e análise de dados de uma forma geral, bem como questões eminentes trazidas pelo potencial impacto da IA Generativa sobre autores, artistas e titulares empresariais, ao mesmo tempo em que pesquisadores, educadores, jornalistas, profissionais das instituições de memória e outros agentes de pesquisa se encontram diante de um vácuo regulatório que compromete a efetivação de seus deveres institucionais. Parte-se de uma visão de Direito Autoral pautada na valorização do interesse público em relação ao acesso ao conhecimento e a seus frutos e do equilíbrio entre a proteção e o acesso ao conteúdo protegido. O trabalho se baseia em pesquisa bibliográfica, documental e empírica, com o uso de técnicas de análise de conteúdo para a coleta, organização e interpretação de entrevistas realizadas com profissionais na área acadêmica, jornalística e de preservação da memória. Também recorre a documentos oficiais, legais e judiciais de diversas regiões ao redor do mundo, com destaque para Brasil, Estados Unidos, Japão, Europa e América Latina. Constata-se que a mineração de textos e dados, ainda que essencial à pesquisa em diversos campos do conhecimento, é dificultada no Brasil por questões relacionadas aos direitos autorais – exacerbadas, mais recentemente, por um processo de recrudescimento do instituto, sob a justificativa de resguardar os interesses de autores, artistas e demais titulares ante a expansão de modelos de IA Generativa. Enquanto isto, a proposta ora contemplada no Congresso se mostra excessivamente restritiva e desconectada das necessidades de profissionais de pesquisa e preservação da memória que dependem da coleta e uso automatizados de materiais. Defende-se, assim, uma ruptura com o paradigma maximalista hoje presente no debate legislativo, em favor de uma visão descentralizada de propriedade, que valorize os múltiplos interesses envolvidos na produção de conhecimento, a diversidade de possibilidades regulatórias e as características e necessidades do ambiente de pesquisa brasileiro.

Palavras-chave: Mineração de dados. Direito Autoral. Pesquisa. Limitações e exceções. Inteligência artificial.

ABSTRACT

Alvarenga, M. B. **O DEBATE REGULATÓRIO SOBRE MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL NA PESQUISA BRASILEIRA.** Orientador: Allan Rocha de Souza. 2025. Tese (Doutorado em Políticas Públicas, Estratégias e Desenvolvimento) – Instituto de Economia – IE, Universidade Federal do Rio de Janeiro, 2025.

This dissertation seeks to evaluate if the current regulatory debate and proposal on the legal regulation of text and data mining has addressed the needs of the research environment, considering public innovation policies and the Brazilian legal system. This research is motivated by the growth in the use of automated data collection and analysis tools in general, as well as eminent issues brought about by the impact of Generative AI on authors, at the same time as researchers, educators, journalists and professionals from museums, libraries and archives, as well as other research agents, find themselves in a regulatory limbo that compromises the effectiveness of their institutional duties. This work adopts a notion of copyright centered around fulfilling the public interest regarding the access to knowledge and to its products. The work is based on bibliographical and documentary research, as well as the use of content analysis techniques to collect, organize and interpret interviews with professionals in the academic, journalistic and memory preservation fields. Official, legal and judicial documents from various regions were also used in the work, with a special focus on Brazil, the United States, Japan, Europe and Latin America. It was found that text and data mining, although essential in various fields of knowledge, is hampered in Brazil by copyright issues, more recently accentuated by a process of increased maximalism in the institute, under the justification of protecting the interests of owners in the face of the expansion of Generative AI models. Meanwhile, the regulation currently under discussion in Congress seems to be excessively restrictive and disconnected from the needs of research and memory preservation professionals who depend on the automated collection and use of materials. Therefore, the work argues for a break from the maximalist paradigm currently present in the legislative debate, in favor of a decentralized vision of ownership, which values the multiple interests involved in the production of knowledge, the diversity of regulatory possibilities and the characteristics and needs of the Brazilian research environment.

Keywords: Data mining. Copyright. Research. Limitations and exceptions. Artificial Intelligence.

LISTA DE ILUSTRAÇÕES

FIGURA 1: Mapa de literatura, contendo as referências bibliográficas principais – p. 28

FIGURA 2: Mapa mental com os principais eixos metodológicos da pesquisa – p. 40

FIGURA 3: Pirâmide DIKW – p. 44

FIGURA 4: Arquitetura de um típico processo de mineração de textos e dados – p. 49

FIGURA 5: Gráfico do Google Trends sobre o uso mundial do termo "Generative AI" em buscas entre setembro de 2021 e janeiro de 2025 – p. 74

FIGURA 6: Mapa com classificação dos países conforme o grau de abertura de Direito Autoral para mineração de textos e dados na pesquisa – p. 108

QUADRO 1: Conjuntos de palavras-chave para definição da *string* de busca – p. 20

QUADRO 2: Resultados da busca no Portal de Periódicos CAPES – p. 21

QUADRO 3: Relação de artigos selecionados após a busca no Portal de Periódicos CAPES – p. 22

QUADRO 4: Resultados da busca em repositórios de teses e dissertações – p. 25

QUADRO 5: Relação de textos selecionados após a busca em repositórios de teses e dissertações – p. 25

QUADRO 6: Ocupações das pessoas entrevistadas selecionadas para o trabalho – p. 169

QUADRO 7: Exemplos de respostas dentro da categoria “uso de dados” – p. 171

QUADRO 8: Exemplos de respostas dentro da categoria “acesso” – p. 172

QUADRO 9: Exemplos de respostas dentro da categoria “autoria e titularidade” – p. 174

QUADRO 10: Exemplos de respostas dentro da categoria “compartilhamento” – p. 176

QUADRO 11: Exemplos de respostas dentro da categoria “dados públicos” – p. 178

QUADRO 12: Exemplos de respostas dentro da categoria “apropriação por terceiros” – p. 180

QUADRO 13: Exemplos de respostas dentro da categoria “clareza regulatória” – p. 182

LISTA DE SIGLAS E ABREVIACÕES

AI – *Artificial Intelligence*

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CDIP – *Committee on Development and Intellectual Property*

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

COPPE – Coordenação dos Programas de Pós-graduação de Engenharia

DIKW – *Data, Information, Knowledge, Wisdom* (Dados, Informação, Conhecimento, Sabedoria)

DRM – *Digital Rights Management* (Gestão de Direitos Digitais)

DVD – *Digital Versatile Disc* (Disco Versátil Digital)

EUA – Estados Unidos da América

GenAI – *Generative AI* (IA Generativa)

IA – Inteligência Artificial

IBGE – Instituto Brasileiro de Geografia e Estatística

IBDAutoral – Instituto Brasileiro de Direitos Autorais

IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia

IoT – *Internet of Things* (Internet das Coisas)

LDA – Lei Brasileira de Direitos Autorais (Lei nº 9.610/98)

LGPD – Lei Geral de Proteção de Dados Pessoais

MPT – Medidas de proteção tecnológica

MTD – Mineração de textos e dados

NIH – National Institute of Health

NoSQL – *No Structured Query Language* (Linguagem de Consulta Não Estruturada)

OMC – Organização Mundial do Comércio

OMPI – Organização Mundial da Propriedade Intelectual

PI – Propriedade intelectual

SCCR - *Standing Committee on Copyright and Related Rights* (Comitê Permanente sobre Direito Autoral e Direitos Conexos)

SINDCINE – Sindicato dos Trabalhadores na Indústria Cinematográfica e do Audiovisual dos Estados de São Paulo, Rio Grande do Sul, Mato Grosso, Mato Grosso do Sul, Goiás, Tocantins e Distrito Federal

SQL – *Structured Query Language* (Linguagem de Consulta Estruturada)

STF – Supremo Tribunal Federal

STJ – Superior Tribunal de Justiça

SSRN – *Social Science Research Network*

TDM – *Text and data mining* (mineração de texto e dados)

TICs – Tecnologias de informação e comunicação

TPM – *Technological Protection Measures* (Medidas Tecnológicas de Proteção)

TRIPS – *(Agreement on) Trade Related Aspects of Intellectual Property Rights* (Acordo sobre Aspectos de Direitos de Propriedade Intelectual Relacionados ao Comércio)

UE – União Europeia

UFRJ – Universidade Federal do Rio de Janeiro

USA – *United States of America*

WIPO – *World Intellectual Property Organization*

WTO – *World Trade Organization*

SUMÁRIO

INTRODUÇÃO	15
1. Contextualização e apresentação do problema.....	15
2. Objetivos.....	27
Objetivo geral	27
Objetivos Específicos	27
3. Referencial teórico	28
3.1. Mineração de textos e dados e bases de dados: usos e aplicações	29
3.2. Instituições de pesquisa e relações universidade-empresa	31
3.3. Proteção e acesso a bancos de dados e seu conteúdo	34
4. Metodologia.....	39
CAPÍTULO 1. MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL	43
1.1. A natureza dos dados e seus repositórios	43
1.2. Mineração de textos e dados	46
1.3. A proteção de bases de dados e seu conteúdo pelo Direito Autoral.....	56
1.3.1. A originalidade nos dados e repositórios.....	59
1.3.2. Mineração de textos e dados e sua licitude no Direito Autoral	63
1.3.3. O sistema de camada tripla no Direito Autoral	68
1.4. O direito <i>sui generis</i>	71
1.5. IA Generativa.....	74
CAPÍTULO 2: LIMITAÇÕES E EXCEÇÕES PARA MINERAÇÃO DE TEXTOS E DADOS NA ESFERA SUPRANACIONAL.....	81
2.1. Conhecimento, regimes de apropriação e Direito Autoral	81
2.2. Uma breve introdução sobre limitações e exceções no Direito Autoral.....	88
2.3. O regime internacional de limitações e exceções	92
2.4. O papel de um uso livre para mineração de textos e dados no atual sistema de limitações e exceções	101
CAPÍTULO 3. LIMITAÇÕES DE DIREITO AUTORAL PARA MINERAÇÃO DE TEXTOS E DADOS NAS EXPERIÊNCIAS NACIONAIS	107

3.1. Estados Unidos: o modelo de cláusula aberta e o uso justo (<i>fair use</i>).....	109
3.2. Os modelos de limitação estatutária para mineração de textos e dados.....	124
3.2.1. Usos livres gerais para pesquisa	125
3.2.2. Japão	127
3.2.3. União Europeia.....	134
3.2.4. O Reino Unido e o sistema “híbrido” do <i>fair dealing</i>	144
3.3. Limitações para mineração de textos e dados na América Latina.....	146
CAPÍTULO 4: MINERAÇÃO DE TEXTOS E DADOS NO ECOSSISTEMA DE PESQUISA	150
4.1. A interseção entre academia e mercado	150
4.2. Ciência aberta e compartilhamento de dados na pesquisa	155
4.3. O compartilhamento de dados por editores e repositórios	162
4.4. A visão do pesquisador.....	166
4.5. Acesso a dados de pesquisa: a visão de profissionais brasileiros	168
4.6. Um ambiente marcado pela complexidade	184
CAPÍTULO 5: OS DEBATES SOBRE MINERAÇÃO DE TEXTOS E DADOS NO DIREITO AUTORAL BRASILEIRO.....	188
5.1. Limitações e exceções no Direito Autoral brasileiro contemporâneo.....	188
5.2. Mineração de textos e dados na política brasileira de inovação.....	195
5.3. Mineração de textos e dados e Direito Autoral no âmbito legislativo.....	196
5.4. Principais questionamentos	209
5.4.1. Escopo dos sistemas de IA a serem regulamentados.....	209
5.4.2. Escopo do uso livre para os agentes de pesquisa.....	210
5.4.3. Questões de transparência	211
5.4.4. Escopo dos atos permitidos	212
5.4.5. A exigência de finalidade não-comercial	213
5.4.6. O uso da regra dos três passos	214
5.4.7. Questões de clareza sobre a exigência de acesso lícito	216
5.4.8. A possibilidade de sobreposição contratual.....	216
5.4.9. Medidas tecnológicas de proteção.....	217

5.4.10. A natureza e aplicação da reserva legal.....	219
5.4.11. Escopo do dever de remuneração	221
5.4.12. Novos direitos exclusivos: quem é o titular?.....	221
5.4.13. O licenciamento como alternativa – como estipular a remuneração?	223
5.4.14. A posição do Estado como mediador	225
5.5. Caindo nas mesmas armadilhas?	226
5.6. Repensando o paradigma regulatório.....	233
CONSIDERAÇÕES FINAIS.....	240
REFERÊNCIAS.....	249
APÊNDICE A – Exemplo de roteiro de entrevista.....	290
APÊNDICE B – Quadro com as emendas sobre mineração de textos e dados ao Projeto de Lei nº 2.338/2023 até 10 de dezembro de 2024	291

INTRODUÇÃO

1. Contextualização e apresentação do problema

Esta tese tem, por objetivo geral, abordar o impacto da proteção aos direitos autorais e seu regime de limitações e exceções sobre o desenvolvimento científico e tecnológico no país, em um contexto no qual a análise computadorizada de grande volume de dados ocupa um protagonismo cada vez maior nas atividades de pesquisa e inovação. Especificamente, busca-se avaliar o estado das discussões em torno da regulação das tecnologias de mineração de textos e dados para fins de pesquisa científica no Brasil, considerando o papel cada vez maior desta na pesquisa intensiva em dados e na inovação tecnológica.

O trabalho se situa em um contexto de crescente valorização do imaterial na economia, que reflete o papel crescente de ativos imateriais na competitividade e na inovação (Julien, 2010, p. 51, 57-59). Neste cenário, a posse, controle e uso de um grande volume de dados passaram a ocupar posição de destaque na atividade econômica e no desenvolvimento científico e tecnológico, sendo utilizadas para uma ampla variedade de fins: do treinamento e operação de sistemas de inteligência artificial e a gestão baseada em dados de governos e empresas, até o desenvolvimento e aprimoramento de bens e serviços, e avanços na realização de pesquisas científicas, que ganham em rapidez e detalhamento (Chen, Chiang e Storey, 2012, p. 1168-1172; McAfee e Brynjolfsson, 2012, p. 6-8; Dean, 2014; Flynn e Vyas, 2022). Exemplo recente foram seus usos na pandemia de COVID-19, essenciais para o sequenciamento genético do SARS-CoV-2 e no desenvolvimento de vacinas. O papel dessas tecnologias, contudo, vai para além da ciência em si, constituindo ferramenta poderosa no jornalismo – a exemplo do jornalismo baseado em dados (Coddington, 2015; Rogers, Schwabish, and Bowers, 2017; Heravi, 2019; Bounegru e Gray, 2021) e do combate à desinformação (Souza, A., Schirru e Alvarenga, 2020) – e nas atividades de projetos e instituições de preservação da memória (Noehrer et al, 2021; Marques, J., 2022), conforme visto em exemplos como o Internet Archive (2024).

A simples posse de dados, contudo, não traz valor por si só: embora revelem diversos atributos a respeito de uma pessoa, objeto ou situações, os dados só têm valor quando contextualizados e relacionados entre si, o que demanda um esforço contínuo de armazenamento e organização (Rowley, 2007; Kelleher e Tierney, 2018). Uma parte crucial deste processo são as técnicas de mineração de textos e dados (também chamadas de *knowledge discovery from data*, *text and data mining*, TDM ou, como propomos neste trabalho, MTD), que permitem extrair padrões e correlações a partir de grandes volumes de dados coletados de diversas fontes (Han, Pei e Kamber, 2011). Tais técnicas são essenciais a pesquisa contemporânea e também

estão no cerne de processos de aprendizado de máquina, ou *machine learning*, que são usados por inteligências artificiais para a execução de suas tarefas (Russel e Norvig, 2013, p. 5; Kelleher e Tierney, 2018; Drexl et al, 2019; Kaplan e Haenlein, 2019).

O crescente protagonismo dos dados e da relevância econômica de repositórios que organizem esse conteúdo levou a uma demanda por regimes institucionais que regulem a apropriação e uso desses dados, a exemplo da ampliação do escopo dos direitos autorais para incluir a proteção aos bancos de dados. A justificativa, aqui, seria proteger os investimentos financeiros, recursos humanos e técnicos depositados na criação e organização desses repositórios (Derclaye, 2008). Contudo, a estrutura de proteção resultante recebeu diversas críticas por estabelecer prerrogativas demais aos titulares e criar uma excessiva restrição para os usuários, criando obstáculos à circulação de informação, à atividade científica e à inovação tecnológica (Reichman e Okediji, 2012; Senftleben, 2016; Hugenholtz, 2017; Banterle, 2018; Levendowski, 2018).

No tocante à mineração de textos e dados, a literatura aponta para a necessidade de regulações, dentro dos direitos autorais, que estabeleçam termos que permitam a coleta e uso de dados para determinados fins – observando que, em muitos casos, e especialmente na pesquisa científica, estas atividades normalmente não interferem com a exploração comercial padrão das obras protegidas por seus respectivos titulares (Gervais, 2008, 2019; Caspers e Guibault, 2016; Senftleben, 2016; European Copyright Society, 2017; Hugenholtz, 2017; Banterle, 2018; Ducato e Stowel, 2018; Geiger, Frosio e Bulayenko, 2018; Alvarenga, 2019; Flynn et al, 2020, 2022b; Carroll, 2020).

Reconhecendo a relevância do tema, algumas jurisdições implementaram limitações de direitos autorais para permitir o uso dessas técnicas, visando promover um melhor ambiente para a inovação tecnológica e científica: exemplos podem ser encontrados nos EUA, no Japão (2004, 2018) na União Europeia (2019) e em Singapura (2021). Propostas similares também são objeto de discussão na América Latina, com graus variados de abertura e implementação (Equador, 2016; Chile, 2024; Uruguai, 2024).

No contexto brasileiro, o Superior Tribunal de Justiça estabeleceu precedente jurídico para um ambiente de mais acolhimento às limitações e exceções no Direito Autoral brasileiro: em 2011, o Recurso Especial nº 964404/ES – posteriormente consolidada no Enunciado 115 da III Jornada de Direito Comercial (Brasil, 2019a) – estabeleceu o caráter exemplificativo do rol de usos livres dos artigos 46 a 48 da Lei de Direito Autoral, permitindo ao Judiciário estabelecer hipóteses de usos que não violem direitos autorais para além daquelas já expressas em lei.

Contudo, conforme apontado em Alvarenga (2019) e em Souza, Schirru e Alvarenga (2020, 2024), isto não anularia a necessidade de uma limitação expressa em lei que comporte a mineração de textos e dados e atividades de pesquisa em um âmbito geral, a fim de reduzir incertezas e dar segurança às atividades nas quais o compartilhamento e uso intensivo de dados ocupam papel cada vez maior.

Nesse sentido, a Estratégia Brasileira de Inteligência Artificial (Brasil, 2021b) estabeleceu a necessidade de discutir uma limitação de direitos autorais específica para a mineração de textos e dados, afirmando a importância dessas técnicas para o treinamento e desenvolvimento da IA e, conseqüentemente, para incentivar a inovação em âmbito nacional e a competitividade de empresas e startups brasileiras em âmbito global. Além disso, Estratégia Nacional de Propriedade Intelectual não só prevê, em seu artigo 1º, o uso da PI enquanto política pública de incentivo ao desenvolvimento e seu equilíbrio com a livre concorrência e interesse social, como, em diversos pontos (a exemplo de seu Eixo 6), prevê o uso da mineração de dados ou da construção de bancos de dados como ferramenta de auxílio do governo ao setor industrial e de tecnologia no Brasil (Brasil, 2021a).

Finalmente, o Projeto de Lei nº 2338/2023 – que hoje aglutina, no Brasil, o debate legislativo sobre IA iniciado na Câmara dos Deputados pelo PL nº 21/2020 e outros projetos subsequentes em ambas as casas do Congresso Nacional (Brasil, 2024g) – traz uma proposta de limitação ao direito de autor para tais atividades, desde que inseridas no contexto específico do treinamento de sistemas de inteligência artificial. Tal proposta foi objeto de discussão no processo de consulta pública do Projeto, contando, inclusive, com a participação de instituições de pesquisa.

Entretanto, uma vez que esta discussão parece haver se iniciado de forma relativamente recente e tendo em vista os recentes desenvolvimentos no campo da IA Generativa, ainda não parece haver clareza sobre como este debate tem se manifestado entre os atores afetados ou de que maneira estes são contemplados no desenho legal da proposta de regulação ora considerada. Indaga-se, assim, sobre como o debate acima é percebido, recebido e discutido pelos próprios profissionais no ambiente acadêmico e outros mais envolvidos com pesquisa intensiva em dados, como no jornalismo e na conservação – especialmente considerando a diversidade de tipos de dados existentes, as diferentes práticas de investigação e apropriação da informação entre comunidades de pesquisadores e a visão das instituições envolvidas sobre este assunto.

Tampouco parece claro como este debate se comunica com o atual arcabouço jurídico nacional e, principalmente, com o papel das instituições de pesquisa nas políticas de inovação

mediante sua integração à atividade econômica, através do fomento às relações universidade-empresa (Etzkowitz e Leydesdorff, 1995, 2000) e à universidade empreendedora (Etzkowitz, 1983; Etzkowitz et al, 2019). Isto inclui considerações sobre o uso da propriedade intelectual como ferramenta de apropriação dos resultados da pesquisa e sua posterior transação como parte do processo de transferência de tecnologia que compõe estas relações (Liu e Van der Sijde, 2021) – o que constitui um tópico já presente na literatura, inclusive no Brasil (dos Santos e Torkomian, 2013; Fischer, Schaeffer e Silveira, 2018; Fischer, Schaeffer e Vonortas, 2020). Contudo, nota-se que não parece claro o papel do Direito Autoral neste processo de apropriação, tampouco como a apropriação de bases de dados e seu conteúdo se dá neste cenário.

Também é verdade que existem trabalhos a respeito do acesso a dados de pesquisa, inclusive no contexto da ciência aberta. Nesse sentido, destacamos Gassman et al (2011), que abordam a ciência aberta como interligada ao processo de inovação aberta, além de Mueller-Langer e Scheufen (2013), que incluem as dificuldades de conciliar ciência aberta com os interesses dos detentores de direitos (tais como editoras de periódicos científicos), Andreoli-Versbach e Mueller-Langer (2013), sobre a baixa inclinação de pesquisadores em divulgar os dados de pesquisa, Machado (2015), sobre a relação entre ciência aberta e dados abertos, e Guanaes (2020), sobre políticas de gestão de dados em periódicos e repositórios abertos. Apesar disto, estas análises não se dedicam a tratar, em alguma medida, da apropriação de bases de dados de pesquisa e seu conteúdo especificamente pelo Direito Autoral, ao menos no ambiente brasileiro.¹

Ainda neste sentido, destaca-se o relatório FORMICT (Brasil, 2023), cujo último estudo observou as políticas de propriedade intelectual de instituições científicas e tecnológicas no Brasil, tomando como referência o ano de 2019. O estudo aponta para uma grande ênfase na gestão de propriedade intelectual, transferência de tecnologia e parcerias com outros atores dentro da política de inovação das instituições de pesquisa analisadas. Não há, contudo, informações específicas acerca da apropriação de dados de pesquisa ou de seus repositórios. Ademais, embora haja menção a dados sobre registros de Direito Autoral, o estudo não explora em maiores detalhes o objeto destes registros – se ele está associado a bases de dados, por exemplo.

Diante deste panorama, optou-se pela realização, em setembro de 2023, de um levantamento sistemático da literatura disponível no Brasil sobre o tema, tomando como base o modelo apresentado por Galvão e Ricarte (2020) e Page et al (2020) – com um segundo

¹ Em âmbito internacional, merece destaque o trabalho de Duro (2019) sobre limitações de Direito Autoral para mineração de textos e dados com a Diretiva 2019/790 da EU e sua aplicabilidade ao contexto brasileiro.

levantamento realizado em janeiro de 2025, a fim de averiguar a existência de eventuais textos relevantes publicados entre a primeira busca e a finalização desta tese. O trabalho buscou diferentes fontes nesse sentido – sendo a primeira delas o repositório disponível no Portal de Periódicos da CAPES. A partir do referencial teórico previamente reunido – e mencionado neste trabalho –, montamos uma estratégia de busca² (ou *string* de pesquisa)³ a partir de palavras-chave relacionadas, de forma direta ou indireta, à interação entre mineração de textos e dados e Direito Autoral. A *string* principal foi construída a partir da combinação das palavras-chave da coluna 1 com as que constam da coluna 2, como pode ser observado na tabela abaixo:

² Ver Galvão e Ricarte (2020).

³ Ver Page et al (2020).

QUADRO 1: Conjuntos de palavras-chave para definição da *string* de busca

<i>Coluna 1</i>	<i>Coluna 2</i>	<i>Coluna 3</i>
Palavra-chave primária	Palavra-chave secundária	<i>String</i> de busca
Direito autoral	<ul style="list-style-type: none"> • Mineração de textos e dados • Mineração de textos • Mineração de dados • Data mining • Descoberta de conhecimento • Análise de dados • Bancos de dados • Bases de dados • Inteligência artificial 	» (“direito autoral” OR “propriedade intelectual” OR “regulação”) AND (“mineração de textos e dados” OR “mineração de dados” OR “mineração de textos” OR “data mining” OR “descoberta de conhecimento” OR “análise de dados” OR “bancos de dados” OR “bases de dados” OR “inteligência artificial”)
Propriedade intelectual	<ul style="list-style-type: none"> • Mineração de textos e dados • Mineração de textos • Mineração de dados • Data mining • Descoberta de conhecimento • Análise de dados • Bancos de dados • Bases de dados • Inteligência artificial 	
Regulação	<ul style="list-style-type: none"> • Mineração de textos e dados • Mineração de textos • Mineração de dados • Data mining • Descoberta de conhecimento • Análise de dados • Bancos de dados • Bases de dados • Inteligência artificial 	

Fonte: Elaborado pelo autor

Foram incluídos os artigos científicos publicados em periódicos nacionais com revisão de pares, contendo estudos primários e publicações escritas em português, e que tenham considerado a interface entre acesso a dados de pesquisa e Direito Autoral como parte do objeto de

pesquisa. Nesta etapa, foram excluídos: revisões de literatura; teses e dissertações; trabalhos que não passaram por revisão por pares; artigos em língua estrangeira e trabalhos que não abordam diretamente a temática de acesso a dados e direitos autorais, bem como editoriais e outros textos sem caráter de pesquisa científica propriamente dita, além de artigos cujo acesso ao texto completo se encontrava restrito.

O levantamento adotou como critério temporal o período entre janeiro de 1991 e janeiro de 2025. A escolha de 1991 como marco inicial se dá pelas transformações ocorridas no ramo da propriedade intelectual com a assinatura de tratados internacionais como o Acordo TRIPS, a Diretiva de Dados da União Europeia e a promulgação da Lei Brasileira de Direitos Autorais na década de 1990 – em especial a inclusão de bancos de dados digitais no conjunto de obras protegidas pelo direito de autor.

Isto posto, realizamos a pesquisa no Portal de Periódicos, chegando aos resultados mostrados abaixo. A coluna da esquerda mostra o total de trabalhos obtidos com a *string* de busca acima mencionada, enquanto a coluna central indica o número de artigos selecionados após a aplicação dos filtros temporais e de exclusão. A coluna da direita indica o número final de artigos relevantes encontrados após a leitura dos resumos e do texto nos trabalhos restantes.

QUADRO 2: Resultados da busca no Portal de Periódicos CAPES

Total de trabalhos encontrados	Trabalhos selecionados (pós-filtro)	Artigos relevantes encontrados
3.977	900	8

Fonte: Elaborado pelo autor

No total, encontramos um total de 8 (oito) artigos, listados em ordem cronológica abaixo:

QUADRO 3: Relação de artigos selecionados após a busca no Portal de Periódicos CAPES

Autor (es)	Obra	Palavras-chave
Ariente, E.A.	Políticas de acesso aberto para trabalhos científicos: interesse público e direitos de autor. Revista brasileira de políticas públicas , 2017, Vol.7 (1)	Direitos autorais. Direito à educação. Acesso Aberto.
Oliveira, J.V.	O FUTURO REPETINDO O PASSADO? DIGITAL RIGHTS MANAGEMENT, TECNOLOGIAS DISRUPTIVAS E O DIREITO AUTORAL BRASILEIRO. Quaestio Iuris , vol. 12, (02), Rio de Janeiro, 2019, pp. 647-672.	Blockchain. Contratos inteligentes. Direito autoral. Inteligência artificial. Gerenciamento de direitos digitais.
Souza, A.; Schirru, L.; Alvarenga, M.	Direitos autorais e mineração de dados e textos no combate à COVID-19 no Brasil. LiinC em Revista , vol. 16 (2), p. 1-15, 2020.	Direito Autoral; Banco de Dados; Mineração de Dados e Textos; Limitações e Exceções; COVID-19.
Rosa, S. S.; Silva, F. C. C.; Pavão, C. M. G.	Iniciativas de acesso aberto no combate à pandemia: dados abertos e propriedade intelectual na disseminação da informação e conhecimento. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação , Campinas, SP, v. 19, n. 00, p. e021025, 2021. DOI: 10.20396/rdbci.v19i00.866688.	Acesso aberto. Dados abertos. Propriedade intelectual. Transparência Covid-19. Open Covid Pledge.
Guanaes, P. C. V.; Albagli, S.	Dados de pesquisa subjacentes a artigos científicos: questões do direito autoral. Em Questão , 2022, Vol.28 (3), p.114171-25.	Dados de pesquisa; Dados de pesquisa subjacentes; Direito autoral; Licenças Creative Commons; Comunicação científica.
Guanaes, P. C. V.; Albagli, S.	Direito Autoral sobre dados de pesquisa no ecossistema da Comunicação Científica. TransInformação , v. 35 I e226918, 2023.	Agências de financiamento. Dados de pesquisa. Direito de autor. Editores. Repositório de dados de pesquisa. Revista científica.

Rezende, L. V. R.; Stueber, K.; Monteiro, E. C. S. A.; Silva, F. C. C. da .; Oliveira, A. F.; Alves, R. A.	Integridade acadêmica no contexto da Ciência Aberta: cenário brasileiro de questões éticas e legais em pesquisas científicas. BiblioCanto , [S. l.], v. 9, n. 2, p. 75 –, 2023. DOI: 10.21680/2447-7842.2023v9n2ID33804. Disponível em: https://periodicos.ufrn.br/bibliocanto/article/view/33804 . Acesso em: 23 jan. 2025.	Integridade acadêmica. Ética na pesquisa. Direitos autorais. Propriedade intelectual. Planos de gestão de dados.
Souza, A.; Schirru, L.; Alvarenga, M.	Mineração de textos e dados na pesquisa em saúde: reflexões sobre direitos autorais. Cadernos de Saúde Pública , v. 40 (5), 2024. DOI: https://doi.org/10.1590/0102-311XPT169023 .	N/A

Fonte: elaborado pelo autor

Nesse sentido, Souza, A.; Schirru e Alvarenga (2020, 2024) destacam a relevância da mineração de dados no contexto das pesquisas em saúde, apontando para uma insuficiência regulatória do Direito Autoral para lidar com dados de pesquisa no presente, o que posicionaria pesquisadores em uma situação vulnerável. Por outro lado, Rosa, Silva e Pavão (2021) destacam a importância das iniciativas de acesso aberto em situações de emergências de saúde, tomando a pandemia de COVID-19 como exemplo. Já Guanaes e Albagli (2022, 2023) abordam os direitos autorais nas políticas de compartilhamento de dados relacionados a trabalhos científicos e sua implementação na prática, por meio de políticas de gestão dos repositórios e regimes de licenciamento.

Os demais artigos não se debruçam diretamente sobre mineração de textos e dados, embora tratem de temas adjacentes, como proteção a bancos de dados e acesso a materiais de pesquisa perante o Direito Autoral: Ariento (2017) analisa as políticas de acesso aberto nos repositórios de diversas instituições de pesquisa no Brasil e no mundo e observa um movimento de maior receptividade ao acesso aberto a trabalhos acadêmicos, mas não menciona dados propriamente ditos. Ainda assim, ressalta a incompatibilidade entre a proibição ao compartilhamento via Direito Autoral e a missão da universidade de propagação do conhecimento. Oliveira (2019, p. 656) brevemente ressalta a importância de repensar o Direito Autoral para permitir o acesso devido a fontes para pesquisa, ecoando o pensamento de Levendowski (2018). Já Rezende et al. (2023) atentam para as práticas de pesquisadores em relação à gestão dos dados e seu compartilhamento no tocante ao Direito Autoral, notando uma significativa ausência de considerações de direitos autorais nos planos de gestão de dados analisados – e, mesmo entre os que mencionavam a questão, havia uma variedade de posturas: de licenças Creative Commons (prevalentes nas Ciências Sociais e Humanas) a embargos ou acesso fechado

(particularmente nas Ciências Biológicas e Engenharias), com uma minoria disponibilizando seus dados como abertos.

Adicionalmente, fez-se uma busca especificamente em repositórios de teses e dissertações, nos meses de setembro de 2023 e em janeiro de 2025, com um protocolo de busca ligeiramente diferente: utilizou-se a *string* (“direito autoral”) AND (“mineração de textos e dados” OR “mineração de dados” OR “mineração de textos” OR “data mining” OR “descoberta de conhecimento” OR “análise de dados” OR “bancos de dados” OR “bases de dados” OR “inteligência artificial”).⁴ Como na busca a periódicos, adotaram-se filtros temporais (janeiro de 1991 a janeiro de 2025) e geográficos (produção nacional), além dos seguintes critérios de exclusão: trabalhos em língua estrangeira, que não abordam diretamente a temática de acesso a dados e direitos autorais ou que não tinham o texto completo disponível para análise.

Como fonte, consultaram-se dois repositórios vinculados ao Governo Federal, que reúnem acervos de instituições de ensino superior públicas e privadas de todo o Brasil: a Biblioteca Digital Brasileira de Teses e Dissertações do IBICT (BDTD) (19 resultados) e o Catálogo de Teses e Dissertações da CAPES (280 resultados).⁵ Também foi efetuada consulta à base de teses e dissertações do Observatório Nacional de Direitos Autorais (ONDA), na sua totalidade (60 resultados). Após a aplicação dos filtros e critérios de exclusão e inclusão, obtiveram-se os seguintes resultados:

⁴ Testes preliminares em 2023 levaram à constatação de que, ao contrário do observado na busca de periódicos, a inserção dos elementos “propriedade intelectual” e “regulação” gerava considerável ruído nos resultados de teses e dissertações nas bases utilizadas (com 1.611 resultados no BDTD e 49.051 no catálogo da CAPES, após o controle temporal). Constatou-se, assim, uma sobreinclusão de trabalhos que não satisfaziam os critérios de inclusão projetados. Isto, mais a necessidade de outras adaptações (como a necessidade de remover a restrição para periódicos revisados por pares), levaram à opção de estabelecer um protocolo novo, distinto do utilizado na busca em periódicos, com o redesenho da *string* de busca nos termos acima expostos.

⁵ Assim como no levantamento realizado no Portal de Periódicos, ao pesquisar por termos no Catálogo da CAPES, recomenda-se digitar a *string* de busca por extenso. Adicionalmente, remover os parênteses ajudou na obtenção de resultados na busca feita em 2025. Na Biblioteca Digital Brasileira de Teses e Dissertações, a *string* funcionou sem problemas na segunda busca, com um simples “copia e cola”.

QUADRO 4: Resultados da busca em repositórios de teses e dissertações

Base utilizada	Trabalhos selecionados (pós-filtro)	Textos relevantes encontrados
Banco Nacional de Teses e Dissertações (BDTD)	19	2
Catálogo de Teses e Dissertações da CAPES	280	2
Observatório Nacional de Direitos Autorais (ONDA)	60	2

Fonte: Elaborado pelo autor

Em todas as três bases, constataram-se apenas dois trabalhos que atendiam aos critérios de inclusão e exclusão. São eles:

QUADRO 5: Relação de textos selecionados após a busca em repositórios de teses e dissertações

Autor, título e ano	Tipo de obra	Palavras-chave
Alvarenga, M. B. Mineração de dados, Big Data e direitos autorais no Brasil . 2019. Dissertação (Mestrado em Políticas Públicas, Estratégias e Desenvolvimento) – Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.	Dissertação	
Guanaes, P. C. V. Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral . 2020. 234f. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.	Tese	

Fonte: Elaborado pelo autor

Neste caso, como no anterior, apenas dois dos textos abordavam a questão de mineração de dados de forma central.⁶ Alvarenga (2019)⁷ aborda a mineração de dados e o Direito Autoral sob a ótica da inovação tecnológica e do regime de limitações e exceções no direito de autor. Já o trabalho de Guanaes (2020), tal como os artigos que lhe sucederam (elencados nos

⁶ Observa-se que, embora tenham sido encontrados trabalhos como os de Gonçalves (2019), Lana, P. (2020) e Schirru (2020), que abordam questões relativas à inteligência artificial sob a ótica do Direito Autoral (especialmente no tocante às obras geradas por IA), optou-se por não incluí-los nos resultados finais, uma vez que a questão dos dados não consta como elemento central do objeto de pesquisa.

⁷ Dissertação deste autor, cuja conclusão serviu como ponto de partida para o presente trabalho.

resultados do Quadro 3), explora as políticas institucionais de compartilhamento de dados científicos na área da saúde, conforme se verá mais adiante.

Assim, constatou-se uma baixa incidência de trabalhos nesses repositórios que abordassem de forma mais específica as interfaces práticas entre mineração de textos e dados e Direito Autoral – e, principalmente, como elas devem ser representadas no desenho de uma regulação em âmbito nacional.⁸ Desta forma, em que pese a existência de alguma literatura sobre Direito Autoral e uso de dados em outros países, e discussões do tema já estejam presentes no debate público, são ainda incipientes no âmbito acadêmico.

Assim, diante desse contexto, enfrentamos a seguinte pergunta de pesquisa: **O debate e proposta regulatórios sobre mineração de textos e dados no Direito Autoral atendem às necessidades do ecossistema de pesquisa, considerando as políticas de inovação e o ordenamento jurídico brasileiro?**

Para responder a esta pergunta, este trabalho se divide em cinco partes: o Capítulo 1 estabelece, em termos gerais, no que consistem a mineração de textos e dados, os bancos de dados e seu conteúdo, sua ligação com a inteligência artificial e como estes elementos técnicos interagem com o Direito Autoral, considerando-se seus princípios fundamentais e o emprego desse sistema na proteção de obras protegidas no ambiente digital, especialmente face aos novos desafios postos pela inteligência artificial. Em seguida, o trabalho se debruça sobre o uso do regime de limitações e exceções ao direito de autor e conexos como ferramenta para permitir a mineração de textos e dados na pesquisa. Nesse sentido, o Capítulo 2 se propõe a apresentar uma discussão teórica sobre a natureza deste instituto enquanto reflexo de tensões entre diferentes regimes de apropriação de bens (sobretudo ligados ao conhecimento) e seu papel, na teoria e na prática, na garantia de proteção do interesse público contra abusos no exercício do Direito Autoral. O Capítulo 3, por sua vez, observa como as L&E vêm sendo mobilizadas para garantir o uso de mineração de textos e dados em diferentes países, com destaque para modelos regulatórios específicos e exemplos de destaque em diferentes regiões do mundo. Feitas estas considerações, parte-se a uma análise de como o ambiente de pesquisa vêm lidando com a apropriação e gestão dos dados de pesquisa, considerando-se as atuais dinâmicas de inovação, o mercado acadêmico e o papel de editores, repositórios e, notadamente, dos próprios pesquisadores. Por fim, o Capítulo 5 se volta para o ordenamento jurídico brasileiro, buscando

⁸ Destaca-se, contudo, o trabalho de Calheiros (2009), que apontou a importância da coleta e análise de informações para empresas de tecnologia no Porto Digital de Pernambuco e os elevados riscos percebidos no uso de tecnologias intensivas em dados diante da ausência de uma regulação nesse sentido.

entender como este vêm considerando a realidade tecnológica, regulatória e científica observada nos capítulos anteriores, especialmente sob a ótica do Projeto de Lei nº 2.338/2023.

Conclui-se este trabalho constatando que o atual debate regulatório sobre a mineração de textos e dados no Brasil recebe forte influência das crescentes preocupações com a expansão de modelos de IA Generativa, manifestando o ressurgimento de uma visão maximalista de Direito Autoral que é, paradoxalmente, reducionista, estando desconectada das complexidades que envolvem o ambiente de pesquisa, o papel deste na política de inovação nacional e as próprias características e preceitos presentes no ordenamento jurídico brasileiro. Produz, assim, uma regulação incapaz de atender devidamente às necessidades de pesquisadores e profissionais de instituições de pesquisa e de preservação da memória. O trabalho conclui, portanto, que o alcance de uma regulação verdadeiramente adequada nesse sentido passa, necessariamente, por uma ruptura com este paradigma maximalista em favor de uma visão de Direito Autoral que escape da noção de propriedade natural, absoluta, monolítica e privada como única forma de gestão possível. Isto requer o reconhecimento da multiplicidade de atores e interesses envolvidos no ambiente de pesquisa, uma abordagem mais crítica dos limites e possibilidades oferecidos pelas diversas opções regulatórias sobre o assunto ao redor do mundo e, acima de tudo, uma maior consciência da interação do Direito Autoral com outros ramos do Direito e seu papel dentro de um cenário mais amplo de fomento ao interesse público.

2. Objetivos

Objetivo geral: Avaliar a adequação do debate e proposta regulatórios sobre mineração de textos e dados às necessidades do ecossistema de pesquisa, considerando as políticas públicas de inovação e o ordenamento jurídico brasileiro.

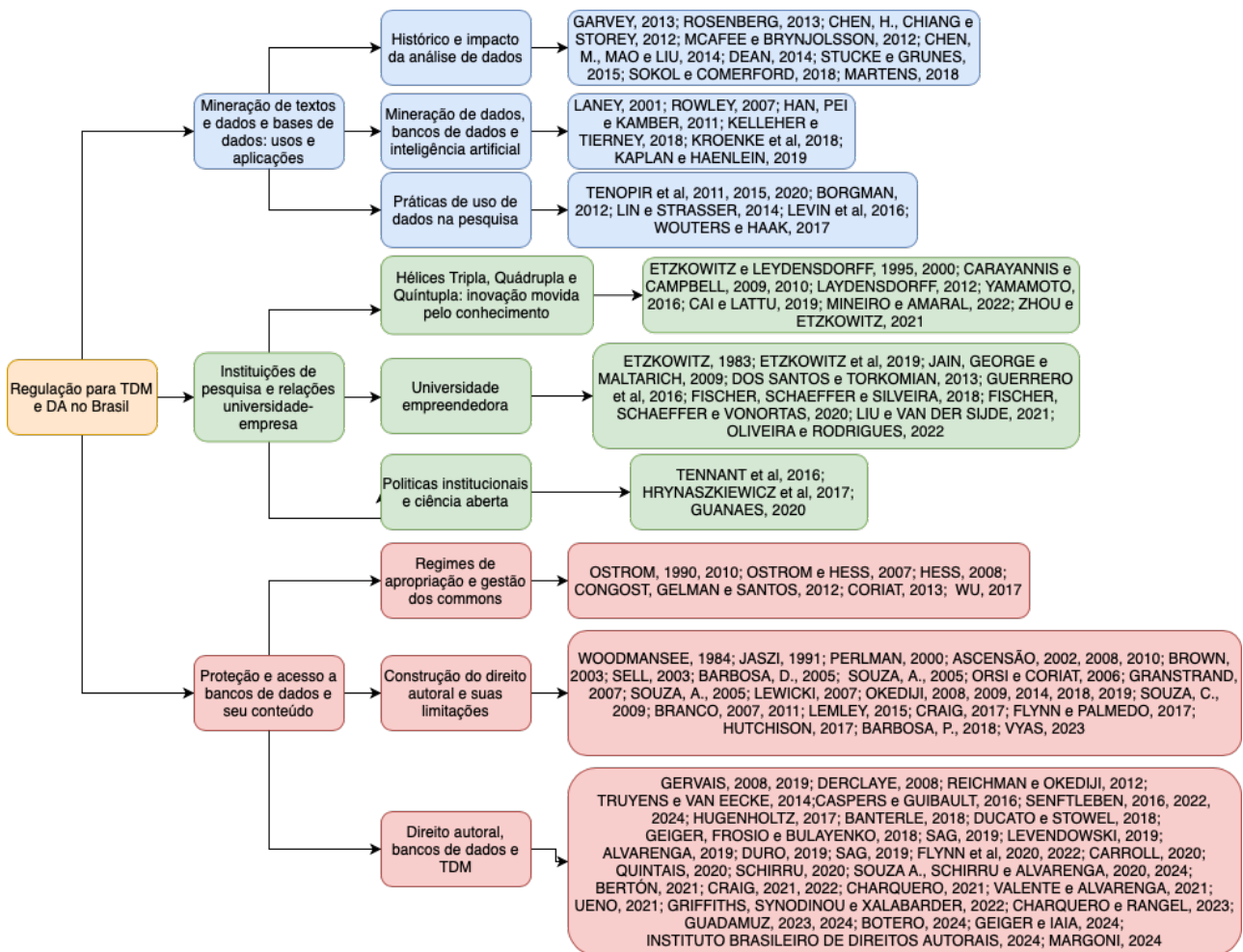
Objetivos Específicos:

1. Investigar os usos e aplicações da mineração de textos e dados no ambiente de pesquisa;
2. Analisar a estrutura e a função do sistema de proteção a bancos de dados e o acesso ao seu conteúdo no Brasil.

3. Referencial teórico

O trabalho ora proposto pretende observar uma questão que repousa na interação entre tecnologia, inovação, direito e política pública. Demanda, portanto, uma interação e cooperação entre disciplinas para enxergar as múltiplas facetas da realidade de uma forma mais completa (Chang, 2012). Isto vai ao encontro da concepção moderna dos estudos de propriedade intelectual, que tentam compreender a PI não mais como fator central para o crescimento econômico, mas como parte de sistemas socioeconômicos mais amplos (Merges, 2017). Com isto em mente, esboçamos o mapa de literatura que consta da Figura 1, inspirado no modelo proposto por Creswell (2007):

Figura 1: Mapa de literatura com as referências bibliográficas principais.



Fonte: elaborado pelo autor

3.1. Mineração de textos e dados e bases de dados: usos e aplicações

A análise do objeto de pesquisa deste trabalho passa, necessariamente, por uma devida compreensão acerca das tecnologias a serem reguladas. Isto envolve discorrer sobre o que se considera como dados e qual o seu valor econômico, tecnológico e social. Ao mesmo tempo, é essencial entender as tecnologias de mineração de textos e dados, seus usos e importância no atual contexto econômico e tecnológico e como os bancos de dados (ou bases de dados, ou *datasets*) integram esse processo. Por fim, olhamos para as práticas de utilização e compartilhamento de dados no ambiente de pesquisa.

Assim, iniciamos com algumas considerações quanto ao conceito de dados, como estes constituem o conhecimento (Rowley, 2007) e como eles se organizam – por exemplo, em dados estruturados e não-estruturados (Kelleher e Tierney, 2018; Kaplan e Haenlein, 2019).

A partir daí, é importante contextualizar os dados e seu valor para a pesquisa como fenômenos que não são recentes. Com efeito, o uso intensivo e sistemático de dados em estudos já foi observado em escritos da Idade Moderna (Rosenberg, 2013), de forma que Garvey (2013) inclusive utiliza o termo “mineração de dados” para descrever o processo de coleta, seleção, organização e publicação de recortes de jornal em um livro sobre a escravidão lançado no Estados Unidos antes da Guerra Civil. Da mesma forma, trabalhar com dados sempre foi um processo custoso – Laney (2001) já falava da existência de dados sendo produzidos em grande velocidade, variedade e volume (os “3 Vs” do Big Data) e, naquele momento, ainda se falava na necessidade de analisar dados por amostragem e obter resultados por aproximação, ao invés de trabalhar com todo o universo possível de informações (Dean, 2014).

O crescimento recente da importância dos dados na atividade econômica e de pesquisa, portanto, deve ser considerado à luz do contexto tecnológico atual. Em outras palavras, a evolução de tecnologias que permitem processar grandes volumes de dados com grande velocidade e especificidade foram os principais responsáveis por um cenário no qual o controle de um grande volume de dados passou a ocupar posição de destaque na atividade econômica e no desenvolvimento científico e tecnológico. Isto porque dados por si só não indicam nada e não carregam nenhum valor imanente – eles são tão valiosos quanto a capacidade de um agente em transformá-los em informação (Rowley, 2007; Dean, 2014).

Aqui se revela a importância da mineração de textos e dados, ou MTD (em inglês, *text and data mining*, ou simplesmente TDM), definida como o processo pelo qual vasculham-se padrões e correlações existentes em grandes volumes de dados, a fim de obter diversas

informações. Tratar-se-ia de uma etapa central em um processo mais amplo de seleção, coleta, agrupamento, organização, limpeza, adaptação e análise de dados (Han, Pei e Kamber, 2011; Truyens e van Eecke, 2014; Kroenke et al, 2016; Kelleher e Tierney, 2018).

Por serem parte essencial do processo de análise de dados, sua presença consequentemente está associada a qualquer atividade intensiva em dados na atualidade, atuando como peça-chave para o desenvolvimento de inúmeras aplicações dentre as quais se destacam: indicação de caminhos ainda inexplorados dentro da pesquisa científica e predição de futuras descobertas; auxílio na gestão de governos e empresas; desenvolvimento de novas tecnologias; aprimoramento de sistemas de segurança e, no campo da saúde, a introdução de melhores diagnósticos médicos e de técnicas de sequenciamento genético para melhor compreender diversas doenças (Chen, Chiang e Storey, 2012, p. 1168-1172; McAfee e Brynjolsson, 2012, p. 6-8).

Uma das principais aplicações da mineração de textos e dados é o treinamento e operação de sistemas de inteligência artificial (IA). Nesse caso, processos de aprendizado de máquina, ou *machine learning*, permitem a uma máquina analisar os dados que lhe são alimentados, interpretá-los e utilizar os resultados desta interpretação como ponto de partida para realizar tarefas e atingir objetivos de forma flexível (Kaplan e Haenlein, 2018, p. 7). Assim, a disponibilidade de grandes quantidades de dados é crucial para que um algoritmo ou IA aprenda e, consequentemente, seja capaz de obter o resultado pretendido de sua operação (Russel e Norvig, 2013, p. 5; Drexler et al, 2019).

De volta à mineração de textos e dados em si, é importante considerar que um componente central para sua realização surge na forma de um registro onde os dados são devidamente organizados, filtrados e preparados com o intuito de facilitar sua preservação e acesso, além de permitir o cruzamento de várias informações com mais eficácia: trata-se do banco (ou base) de dados digital (Kroenke et al, 2016).

Os efeitos da valorização de grandes bancos de dados e a preocupação com o controle sobre os mesmos são sentidos na atividade econômica. Neste sentido, dois pontos de vista divergentes podem ser percebidos: por um lado, autores como Stucke e Grunes (2015) demonstram preocupação com um movimento crescente de diversas empresas no sentido de capturar grandes bancos de dados e torná-los exclusivos – seja pela aquisição e fusão de empresas, seja pela imposição de barreiras ao acesso e uso. Por outro lado, Sokol e Comerford (2016) entendem que dados teriam valor econômico rapidamente decadente (repositórios necessitam ser constantemente atualizados), seriam difíceis de conter ou de ter sua reprodução restringida, permitindo uma baixa barreira à criação de novos serviços ou bancos de dados mais acessíveis.

Finalmente, situamos também nesta seção a observação das práticas acadêmicas em relação à apropriação, uso e compartilhamento de dados por meio do Direito Autoral na atividade de pesquisa. Mais especificamente, tomamos como referência os estudos conduzidos por Tenopir et al (2011, 2015, 2020) a respeito das práticas e visões de pesquisadores acerca do compartilhamento de dados científicos, bem como fatores que surgem como obstáculos e facilitadores desse processo – questão que viria a ser também abordada em trabalhos como os de Levin et al (2016), nas áreas de biologia de sistemas, biologia sintética e bioinformática, e Wouters e Haak (2017), sobre a percepção dos pesquisadores sobre a titularidade dos dados de pesquisa e o seu compartilhamento. Já Borgman (2012) chama atenção para as dificuldades oriundas da grande diversidade de posturas em relação à abertura de dados de pesquisa quando consideramos comunidades científicas diferentes. Lin e Strasser (2014), por sua vez, abordam os intermediários e grandes titulares na divulgação científica em uma consulta a diversos pesquisadores sobre o papel dos editores de periódicos na promoção do acesso e disponibilidade de dados.

3.2. Instituições de pesquisa e relações universidade-empresa

Abordar o tema da regulação sobre tecnologias de interesse para atividades de pesquisa implica, necessariamente, na necessidade de compreender como esta dinâmica interage com a inovação científica e tecnológica e, mais especificamente, com as instituições de pesquisa que são partícipes desse processo.

Para tanto, este trabalho parte da visão de inovação como um processo sistêmico, tomando como referência o modelo de Tripla Hélice proposto por Etzkowitz e Leydesdorff, o qual busca afastar a noção de inovação meramente impulsionada pelas empresas ou provocada pelos usuários, ao mesmo tempo em que visa superar modelos antigos de ecossistemas de inovação como os Modos 1 (soviético, onde o Estado engloba todo o processo de inovação) e 2 (antigo sistema ocidental, apoiado na serendipidade e no *laissez-faire*), adotando o Modo 3 – um sistema de geração de conhecimento onde universidade, empresas e Estado operam em rede, com pontos de interseção. Isto implica na transformação da universidade em um ente não apenas dedicado ao ensino, mas também à pesquisa e à extensão (Etzkowitz e Leydesdorff, 1995; 2000). Nesse sentido, a Tripla Hélice coloca o processo de inovação como centrado no conhecimento e, portanto, as universidades e instituições de pesquisa como atores de destaque – em contraste com o Sistema Nacional de Inovação, centrado na firma, e com o Triângulo de Sábado, voltado para o Estado (Etzkowitz e Leydesdorff, 2000).

O trabalho também toma como base a revisão bibliográfica sobre a evolução da Tripla Hélice e derivações conduzida por Mineiro e Amaral (2022) e reconhece os modelos derivados da Tripla Hélice – que incluíram a demanda pela inserção de outros campos da sociedade ao modelo, tal como visto em Ferreira e Farinha (2013), quando estes tratam do desenvolvimento regional e incluem fatores socioculturais, ambientais, políticos e econômicos. A isto se sucedeu o modelo de Quádrupla Hélice, que visualiza uma estrutura de hélices a nível local e internacional, comportando múltiplos clusters de conhecimento coexistentes em um mesmo sistema de inovação, a exemplo de comunidades científicas diferentes com sistemas diferentes de produção de conhecimento, interação com empresas e formas de enxergar a apropriação e uso de dados (Carayannis e Campbell, 2009; Cai e Lattu, 2019), e que finalmente inclui a sociedade civil como a quarta hélice – e a base institucional para o funcionamento das outras três esferas. A este mecanismo, pode ser adicionada um componente adicional: a influência do meio ambiente e a necessidade de adaptação das políticas de inovação a um modelo de desenvolvimento sustentável. Cria-se, assim, uma Quintupla Hélice (Carayannis e Campbell, 2010, 2011).

A Tripla Hélice ainda passaria por diversas revisões além destas – de fato, Leydesdorff (2012) criou a noção de infinitas hélices como forma de destacar que não há limitação à quantos (ou quais atores inserir dentro do modelo conceitual de hélices, desde que contribuam para a produção de conhecimento. Assim, outros modelos para além da 3/4/5H viriam a surgir, tais como as 6 Hélices de Yamamoto (2016), voltado para parques tecnológicos sustentáveis, e as Triplas Hélices Gêmeas (Zhou e Etzkowitz, 2021).

Um elemento crucial de todos os modelos acima é a universidade empreendedora, observada por Etzkowitz (1983) ao notar um movimento maior de pesquisadores estadunidenses em desenvolver modelos de negócio a partir de suas pesquisas, bem como uma tendência maior entre as universidades no sentido de buscar novas fontes de financiamento a partir de parcerias com empresas. Assim, a universidade passa a executar uma nova missão, saindo da “torre de marfim” e deslocando-se para um papel de maior interação com outros segmentos da sociedade (Etzkowitz et al, 2019). Para Liu e Van der Sijde (2021), essa missão contempla uma série de atividades, não se restringindo à formação de profissionais qualificados para o mercado, mas também mantém a produção científica, porém atrelada a outras atividades, tais como: proteção dos resultados da pesquisa por meio da propriedade intelectual, criação de empresas com base nessa produção (*spin-offs*) e atuação de escritórios de transferência de tecnologia.

Por outro lado, Jain, George e Maltarich (2009) observaram o comportamento de pesquisadores e observaram que, mesmo havendo uma maior comunicação com a indústria, os

acadêmicos entrevistados mantinham uma forte prioridade na pesquisa – e há uma forte preocupação em usar a apropriação dos resultados de pesquisa como forma de trazer mais ganhos financeiros que impulsionarão futuras pesquisas e como forma de prevenir que terceiros utilizem esses resultados de forma abusiva ou antiética. Já Guerrero et al (2016) apontam para uma extensa literatura que aborda o tema a partir de diversos pontos de vista: seja ele a nível individual, lidando com o comportamento de estudantes e pesquisadores, seja a nível organizacional, observando as interações com outros atores do ecossistema de inovação e o funcionamento de processos de transferência de tecnologia. Oliveira e Rodrigues (2022), por sua vez, observam que a transformação não ocorre de forma homogênea nas universidades: existem modelos variados de universidade empreendedora e de ações comumente tomadas nesse sentido: há instituições que focam mais no estímulo à pesquisa colaborativa e geração de propriedade intelectual, enquanto outras se concentram sobre a consultoria. O estudo ainda aponta para graus diferentes de integração dessas atividades com a missão da universidade de disseminação do conhecimento e a geração de benefícios não-pecuniários, além da recompensa financeira.

Finalmente, é importante destacar o ambiente brasileiro, destacando de que forma a universidade tem se integrado com os outros agentes da Tripla Hélice. Aqui, citamos o trabalho de dos Santos e Torkomian (2013), que analisam a atuação de escritórios de transferência de tecnologia em universidades brasileiras, e os levantamentos feitos por Fischer, Schaeffer e Silveira (2018), que analisaram o impacto das universidades na tração de investimentos em inovação a nível regional, e por Fischer, Schaeffer e Vonortas (2020), sobre a evolução do patenteamento feito por universidades e seu nível de interação com diferentes setores industriais, ambos visando medir o grau de integração da academia ao sistema de inovação brasileiro. Observa-se que, no Brasil, este movimento ainda não se encontra em um estado suficientemente avançado de participação das universidades, de forma que ainda é necessário investir em uma atuação mais enfática por parte das instituições de pesquisa.

Por fim, em razão do próprio caráter sistêmico da inovação, é importante considerar a forma como as próprias instituições de pesquisa (e os entes com os quais interagem) instituem processos e políticas internas em relação ao uso de dados. Mais especificamente, voltamos para políticas de abertura de dados na ciência, tratadas por Tennant et al (2016) de forma mais ampla, destacando seus impactos, benefícios e desafios. Hrynaszkiewicz et al (2017), por sua vez, observam diferentes periódicos e identificam quatro tipos de política de compartilhamento de dados científicos. Por fim, destacamos o trabalho de Guanaes (2020) sobre políticas institucionais de compartilhamento de dados científicos na área da saúde – no qual se observa que a

temática da abertura de dados de pesquisa e do papel do Direito Autoral neste cenário parece ainda figurar em segundo plano nas instituições observadas, situando-se em uma situação paradoxal: embora a abertura de dados seja apoiada por agentes importantes dentro da academia, ainda não se verifica um ambiente institucional e regulatório que traga uma efetiva operacionalização dessa abertura, tampouco forneça incentivo à conservação e divulgação de dados de pesquisa ou garanta segurança aos pesquisadores que precisam utilizar material de terceiros.

3.3. Proteção e acesso a bancos de dados e seu conteúdo

Por fim, é crucial entender de que forma essas políticas se materializam em um sistema de regulação sobre a proteção aos bancos de dados e as formas pelas quais o seu conteúdo será acessado e utilizado. Mais especificamente, olhamos para a estrutura desse sistema e sua função dentro de um contexto de estímulo ao desenvolvimento científico, tecnológico e cultural.

Um elemento-chave nessa discussão é a questão dos regimes de apropriação, os direitos de propriedade e a gestão dos recursos comuns, ou *commons* – aqui entendidos como aqueles bens, materiais ou imateriais, que são compartilhados por um grupo de pessoas, vulneráveis a cercamentos (restrições ao uso), exploração excessiva ou dilemas sociais e que, diferentemente de bens públicos, requerem esforços de proteção e administração para se manterem (Hess, 2008).

Aqui, destacamos o papel dos estudos dedicados à gestão coletiva da propriedade e sua crítica à dicotomia público/privado da apropriação dos bens. Ostrom (1990), por exemplo, aponta para a existência de diversos modelos possíveis de gestão, adequados a tipos distintos de bens. Nesse sentido, a autora se opõe a uma visão limitada à esfera puramente pública ou privada, chamando atenção para a possibilidade de um regime de gestão coletiva dos *commons*, na qual se “distribuem” prerrogativas, poderes e responsabilidades entre os usufrutuários daquele bem conforme for mais apropriado para a sua manutenção.

Isto seria possível pois não haveria um único direito de propriedade, mas um feixe de direitos de propriedade (ou *bundle of rights*): *acesso, extração, gerenciamento, exclusão e alienação (e contribuir ou remover, no caso do ambiente digital)*. Diferentes indivíduos podem possuir diferentes tipos de direitos sobre o bem, e diferentes modelos de apropriação (privado, estatal e comunal) podem ser aplicados de forma conjunta, formando sistemas híbridos: por exemplo, parte de um mesmo patrimônio pode ser controlada de forma privada, enquanto outra é compartilhada e gerida coletivamente (Ostrom e Hess, 2007a). Tais noções são importantes

quando consideramos a relação entre a proteção legal conferida a bases de dados e a que é estendida a seu conteúdo: institutos separados, mas diretamente relacionados.

Coriat (2013) avança nesta discussão e se debruça sobre a gestão dos commons centrados em bens imateriais e outros associados a conhecimento (como, por exemplo, repositórios de dados), aqui chamados de *knowledge commons*. O autor chama atenção para o fato de esses recursos, por não se degradarem com o uso, serem geridos não com o intuito de garantir sua preservação, mas sim seu crescimento, e para a necessidade de protegê-los contra a apropriação indevida, que reduz o patrimônio disponível aos que dele desfrutam.

No campo mais específico da propriedade intelectual, chamamos atenção para as fundações do direito de autor com o trabalho de Woodmansee (1984) sobre a construção do Direito Autoral e da noção do trabalho artístico como dotado de valor abstrato, justificando a valorização dos escritores como categoria profissional a ser protegida diante das mudanças na forma de comercializar obras, enquanto Jaszi (1991), denuncia o uso da ideia de “autor” e “obra” como elementos retóricos maleáveis para justificar ampliações e reconfigurações no direito de autor, utilizando-se do argumento de proteção aos interesses do autor e a elevação de diversas obras à posição de obra artística e literária como uma “fachada” para a simples proteção de investimentos.

A culminação desse processo pode ser vista em um aparente deslocamento da regulação de Direito Autoral para a Organização Mundial do Comércio (OMC) na década de 1990, com a assinatura do Acordo TRIPS e a consolidação de esforços nacionais e internacionais de “expansionismo” da PI e “cerceamento” dos bens científicos por parte de países líderes em tecnologias intensivas em informação. Autores como Sell (2003), Orsi e Coriat (2006) e Okediji (2008, 2009) apontam que este movimento surgiu como estratégia para barrar concorrentes e imitadores, desconsiderando as mudanças na forma como autores, titulares e usuários viriam a interagir com as obras protegidas graças às tecnologias da informação. No campo do Direito Autoral, isto envolveu a inclusão de bases de dados e programas de computador ao escopo de materiais protegidos pelo instituto – algo que Ascensão (2002, 2008) critica como um sinal de mercantilização e desvirtuação do instituto, agravado por um processo de apagamento do regime de limitações e exceções, que deveriam ser igualmente revistas para acompanhar esta expansão.

Esta aparente “desvirtuação” nos convida a reanalisar certos aspectos do direito de autor. Aqui, merecem atenção as considerações de Barbosa, D. (2005) sobre a aplicação do requisito da originalidade a programas de computador – e como o caráter subjetivo de

originalidade vai cedendo vez a uma conceituação mais objetiva, de não-banalidade. Outros trabalhos viriam a explorar este instituto por outros pontos de vista, a exemplo da comparação da noção de criatividade na psicologia e no direito de autor (Hutchison, 2017) e da aferição do que seria “original” em um cenário de massificação cultural, onde produtos cada vez mais genéricos e “lugar-comum” tendem a ser mais circulados (Barbosa, P. 2018).

Também fazemos uso do modelo proposto por Dusollier (2005), que considera um sistema de proteção a estrutura de três camadas da propriedade intelectual. Seriam elas: a proteção explícita a bancos de dados e a seu conteúdo, seja pelo direito autoral ou pelo direito *sui generis* europeu; a adoção de medidas tecnológicas de proteção (MTP) e de ferramentas de gestão de direitos digitais (*Digital Rights Management*, ou DRM) e, por fim, as “medidas anti-burla” (do inglês *anti-circumvention measures*), que visam proibir práticas de invalidação das MTP/DRM. Esta última camada, sua natureza peculiar dentro do sistema de Direito Autoral e seu impacto sobre o mesmo já haviam sido observados mais atentamente por Brown (2003). Derclaye (2008) explora como a proteção aos bancos de dados se dá através de diferentes institutos jurídicos ao longo dessas camadas, discutindo as características, justificativas, vantagens e desvantagens de cada uma perante os direitos dos titulares e o interesse público.

Em outra vertente desta seção, observa-se o crescimento de uma literatura favorável a uma maior abertura no Direito Autoral e a um resgate das limitações e exceções como instrumentos de garantia da função social da propriedade e da harmonização do Direito Autoral com outros direitos fundamentais (Souza, A., 2005; Lewicki, 2007; Souza, C., 2009; Ascensão, 2010; Wachowicz, 2010; Branco, 2011). Mais recentemente, em âmbito internacional, chama atenção o trabalho de Flynn e Palmedo (2017) sobre a relação positiva entre regimes de direitos autorais mais abertos em favor dos usuários e o desempenho econômico e acadêmico de um país, além do descompasso entre países do Norte e do Sul Global nesse sentido.

Contudo, também é pertinente que se faça uma leitura mais crítica do próprio sistema internacional de limitações e exceções em si, a exemplo de autores como Perlman (2000), Okediji (2014; 2018), Lemley (2015), Craig (2017) e Vyas (2023), que alertam para a necessidade de se observar os limites políticos (como pressões internacionais e ambiente institucional interno) existentes no atual modelo de L&E de Berna/TRIPS. Mais especificamente, aponta-se que, embora o uso das limitações e exceções como ferramenta de balanceamento e promoção do interesse público seja inestimável, este argumento permanece preso a um regime filosófico e político que supervaloriza o Direito Autoral em detrimento dos demais direitos fundamentais,

cuja observação acrítica enfraquece mobilizações para desenhos institucionais que verdadeiramente observem o interesse público.

Outra crítica, mais específica ao caso da União Europeia, pode ser vista no trabalho de Quintais (2020) a respeito dos dispositivos da Diretiva 2019/790 do Mercado Único Europeu. O autor nota que, embora haja diversas novas disposições no sentido de adequar as limitações e exceções ao ambiente digital – o que inclui artigos que tratam especificamente da mineração de textos e dados –, estas possuem diversos pontos fracos que comprometem sua efetiva aplicação. No geral, o autor aponta que a versão final da Diretiva, terminou refletindo muito mais o *lobby* de grandes titulares de direitos do que uma discussão verdadeiramente pautada em evidências empíricas e conhecimento de *experts*. Conclusões parecidas podem ser vistas em Griffiths, Synodinou e Xalabarder (2022) e Margoni (2024).

No que tange especificamente à mineração de textos e dados, os últimos anos viram o crescimento, a nível internacional, de uma literatura voltada para a necessidade de maior segurança para essas atividades, no campo do Direito Autoral. Hugenholtz (2017) questiona a existência, em si, de um direito de propriedade específico para dados, ao apontar que tal proteção contribuiria para reduzir a acessibilidade de dados públicos e o papel da originalidade, criar impedimentos indevidos aos direitos de usuários, à livre iniciativa e à liberdade de informação, e gerar um clima de incerteza regulatória para aqueles que trabalham com dados. Este último ponto é aprofundado por Geiger, Frosio e Bulayenko (2018), quando estes discutem as maneiras pelas quais o Direito Autoral interfere na coleta e análise de dados.

Mais recentemente, outros textos foram além, colocando a mineração de textos e dados como atividade que, por sua própria natureza, sequer poderia ser considerada ilícita no âmbito do Direito Autoral. Este argumento se baseia principalmente na noção de que tais técnicas e procedimentos não fazem uso da expressão artística ou literária – que é o verdadeiro objeto de proteção do instituto – e, conseqüentemente, não interferem com os direitos exclusivos garantidos aos titulares das obras usadas nesses processos tecnológicos (European Copyright Society, 2017; Sag, 2019; Carroll, 2020; Craig, 2021b), embora isto não necessariamente afaste a observação de elementos fáticos que indicariam o emprego de atos que, em tese, ensejariam a interferência com direitos exclusivos de titulares (Geiger, Frosio e Bulayenko, 2018; Craig, 2021a).

A evolução dos sistemas de IA, contudo, também chama atenção para como modelos processam conteúdo: Sobel (2017) aponta que diversos sistemas são cada vez mais capazes de coletar e processar conteúdo em volume e granularidade suficientes para capturar não só dados não-originais, mas as combinações entre esses dados que podem constituir material original.

Embora certos autores (Geiger, 2022; Senftleben, 2023; Geiger e Iaia, 2024) apontem esta crescente sofisticação da IA como possível justificativa para um maior reforço na proteção dos direitos dos titulares, Guadamuz (2023, 2024), por outro lado, aponta para as dificuldades técnicas em auferir estas observações, uma vez que modelos de IA não processam nem armazenam dados de maneiras necessariamente inteligíveis para seres humanos.

Ao mesmo tempo, Senftleben, que já havia apontado para os efeitos deletérios do atual sistema de proteção a dados via Direito Autoral enquanto barreira para a concorrência (2016), traz complicações para a discussão da mineração de dados no âmbito dos tratados internacionais. Baseando-se em diversos embates diplomáticos envolvendo o estabelecimento da regra dos três passos e a ampliação do escopo dos direitos de reprodução, o autor conclui que, por ora, não há como propor uma regulação internacional para a coleta e análise de dados sem que antes haja consenso sobre esta atividade constituir um ato de reprodução no âmbito da Convenção de Berna (Senftleben, 2022).

Outros autores apontam para problemas decorrentes do uso da propriedade intelectual para a pesquisa científica quando esta se estende para os dados de pesquisa. Reichman e Okediji (2012) já abordavam esse processo de expansão de uma forma mais ampla, incluindo em sua discussão o campo das bases de dados, as barreiras para pesquisadores que desejam lidar com técnicas automatizadas de “descobrimto de conhecimento” e o papel de periódicos e financiadores nesse processo. Ducato e Stowel (2018) apontam como restrições na propriedade intelectual relativas ao uso de dados vêm operando nas relações contratuais em plataformas digitais e como estes interferem em certas tecnologias benéficas ao usuário (a exemplo de programas que simplificam os termos de uso para leigos). Levendowski (2018), por outro lado, reforça que a preferência por uso de repositórios “de baixo risco” para o treinamento de algoritmos tende a potencializar a exacerbação de resultados enviesados em sistemas de IA, em razão da baixa quantidade ou qualidade dos dados utilizados.

Em um campo regulatório mais amplo, Flynn et al (2022a) partem da discussão de Flynn e Palmedo (2017) para se debruçarem sobre mineração de dados e direito à pesquisa. Ali, os autores reiteram a observação de um abismo regulatório entre países mais economicamente desenvolvidos e nações em desenvolvimento: estes últimos apresentaram uma regulação de PI muito mais restritiva para técnicas de coleta e análise de dados ou para procedimentos necessários à pesquisa científica. Nesse sentido, diversos pesquisadores da área já se manifestaram publicamente sobre a necessidade de se reconhecer os direitos dos usuários no campo da pesquisa, inclusive no que diz respeito à inteligência artificial (Caspers e Guibault, 2016; Ducato

e Strowel, 2018; Flynn et al, 2022b) Na América Latina, também se coloca a importância deste tipo de regulação como forma de reduzir o distanciamento tecnológico da região em relação ao Norte Global (Charquero e Rangel, 2023; Botero, 2024).

Por fim, retomamos Ducato e Stowel (2018) e tomamos como base os trabalhos de Duro (2019), Alvarenga (2019), Schirru (2020), Souza A., Schirru e Alvarenga (2020, 2024), Charquero (2021) e Ueno (2021) a respeito da importância da mineração de textos e dados para campos como a ciência e o treinamento de inteligências artificiais, bem como a instituição de limitações para mineração de textos e dados em diferentes contextos jurídicos, trazendo exemplos do Japão, União Europeia, EUA e Equador – e discutindo como o ambiente jurídico do Brasil poderia dar suporte a esse tipo de dispositivo e como o país vem abordando a questão (Valente e Alvarenga, 2021; Schirru et al, 2024, Instituto Brasileiro de Direitos Autorais, 2024).

4. Metodologia

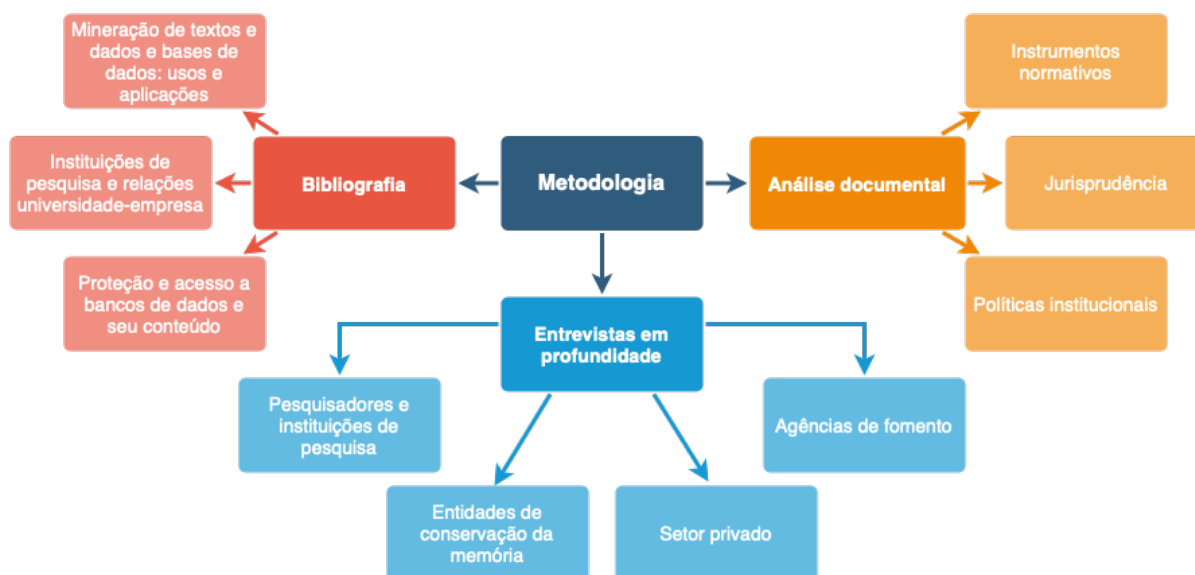
O objetivo da pesquisa ora proposta é avaliar a adequação do debate e proposta regulatórios sobre a mineração de textos e dados ao ecossistema de pesquisa no Brasil, tendo em mente o processo de formulação das atuais políticas públicas de inovação que tocam o tema, bem como o contexto do ordenamento jurídico brasileiro no presente. Isto envolve, necessariamente, a compreensão desse ambiente institucional, seu processo de construção e os interesses e finalidades que o informam. Nesse sentido, embora a pesquisa científica se coloque em posição de destaque, em virtude de sua presença na literatura sobre inovação (conforme visto acima), pontua-se, também, a importância desta regulação para atividades de pesquisa fora da academia, como no jornalismo (Coddington, 2015; Rogers, Schwabish, and Bowers, 2017; Heravi, 2019; Souza, A., Schirru e Alvarenga, 2020; Bounegru e Gray, 2021) e no trabalho de conservação e gestão de coleções (Noehrer et al, 2021; Marques, J., 2022).

Para tanto, optou-se por fazer uso de técnicas de pesquisa qualitativa, descritas por Creswell (2007) como aquelas em que as alegações de conhecimento se baseiam sobretudo em desenvolver uma teoria ou um padrão a partir de significados social e historicamente construídos extraídos de experiências individuais, e que buscam enxergar os fenômenos sociais de forma holística, com múltiplas estratégias de investigação. A pesquisa qualitativa também se caracteriza pelo emprego de técnicas de pesquisa de investigação como narrativas, fenomenologias, etnografias, estudos baseados em teoria ou estudos de teoria embasada na realidade e dá preferência à coleta de dados emergentes abertos, a partir dos quais se busca desenvolver o tema de escolha.

Quanto ao objeto, embora reconheça-se que as atividades de coleta e análise de dados abrangem uma multitude de aplicações e finalidades de relevância para a inovação, escolhemos dar foco ao uso dessas técnicas para pesquisa por diversas razões. A primeira é que este tópico em particular se encontra em um estado mais avançado de discussão a nível mundial, sendo contemplado no ordenamento jurídico de diversos países – o que viabiliza uma observação mais profunda e completa dos diversos modelos regulatórios existentes sobre a mineração de textos e dados. A segunda razão é que, apesar disso, não há um modelo estabelecido ou hegemônico de regulação sobre o tema, o que nos convida a observar como este instituto seria adaptado para o contexto brasileiro. Por fim, este tópico também levanta questões que possuem grande relevância para a regulação da mineração de textos e dados para outros fins. Exemplos são a distinção entre fins comerciais e não-comerciais, as práticas dos usuários e agentes envolvidos, entre outros.

Em relação aos métodos, a fim de permitir a triangulação dos dados e obter uma visão mais completa acerca do nosso objeto, a pesquisa se dividirá em três eixos principais, representados na Figura 2, que estarão presentes ao longo de todo o trabalho:

Figura 2: Mapa mental com os três principais eixos metodológicos da pesquisa



Fonte: elaborado pelo autor, 2025

O primeiro eixo envolve o uso de pesquisa bibliográfica, levando em consideração aspectos de diversos campos do conhecimento relacionados à discussão ora proposta: mineração de textos e dados, bancos de dados, inteligência artificial, direitos autorais e inovação. Além da consulta às principais referências a esses temas, também pretende-se utilizar publicações obtidas por meio de múltiplos bancos de dados, como o Portal de Periódicos da CAPES, o

SSRN, o Google Scholar e os bancos de teses e dissertações pertencentes a universidades públicas e privadas no Brasil.

O segundo eixo se baseia em análise de documentos jurídicos e políticos relacionados à propriedade intelectual e a políticas de inovação referentes a bancos de dados, software e mineração de dados. Em um âmbito global, buscaremos tratados internacionais, como a Convenção de Berna sobre Direito do Autor e Direitos Conexos e o Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (o Acordo TRIPS). Na esfera nacional e supranacional, atentamos para diretivas, legislações, decisões judiciais e documentos oficiais de diferentes jurisdições ao redor do mundo que tenham incorporado questões relacionadas à coleta, análise e gestão de dados em suas regulações sobre inovação e propriedade intelectual. Outra parte do eixo diz respeito a políticas de dados, termos de uso e outros documentos utilizados por instituições de pesquisa no Brasil referentes à apropriação, uso, divulgação e compartilhamento dos dados coletados e produzidos no âmbito da investigação científica.

O terceiro eixo foca na aplicação de entrevistas individuais em profundidade (Gaskell, 2002) com o propósito de buscar diferentes pontos de vista acerca de uma determinada questão – nesse caso, o uso de dados na ciência, sua apropriação e problemáticas decorrentes da propriedade intelectual sobre esse material. Nesse sentido, tomamos como base a entrevista centrada no problema, pautada na orientação para um problema relevante e para um objeto específico de pesquisa e com orientação ao processo. Em outras palavras, utiliza-se um roteiro que sirva de base para a coleta de dados biográficos referentes a uma questão em particular. No caso do trabalho ora proposto, o objetivo é observar como o Direito Autoral influencia nas práticas de pesquisa e uso de dados pelos atores selecionados (Flick, 2009).

Em relação aos entrevistados, foram selecionados pesquisadores (preferencialmente representantes de laboratórios e equipes de pesquisa científica e jornalística), bem como representantes de instituições de pesquisa, agências de fomento, entidades de conservação e setor privado (incluindo jornalismo), visando profissionais que lidem de forma mais próxima com o uso de dados na atividade de pesquisa e as eventuais questões de propriedade intelectual daí resultantes. Atendendo às preocupações de Gaskell (2002, p. 70-71) quanto à viabilidade e necessidade de *corpus* excessivamente grandes de entrevistas, contatou-se 31 potenciais pessoas entrevistadas, das quais nove (9) foram selecionadas para a análise final. O material utilizado compreende um conjunto de entrevistas realizadas com diversos profissionais de pesquisa no Brasil, como parte de um estudo mais amplo conduzido pelo Projeto Arcadia envolvendo experiências com práticas de pesquisa intensiva e, dados e propriedade intelectual na América

Latina, realizado entre agosto e novembro de 2021 (Valente, Lana, A. e Houang, 2024, no prelo), do qual o presente trabalho constitui um desdobramento.

Buscou-se manter o anonimato em relação aos nomes de todas as pessoas entrevistadas, utilizando-se uma linguagem neutra em referência ao gênero sempre que possível – com tratamento padronizado para o feminino nos demais casos – e uma identificação numérica, sorteada de forma aleatória. Assim, cada pessoa será identificada neste trabalho como “Pessoa Entrevistada [número]”. Quanto à ocupação, qualificamos as respondentes apenas na medida necessária para situar sua posição profissional, evitando menções explícitas a instituições ou regiões específicas. Ex.: “pesquisadora”, “representante de agência de fomento”.

Como forma de auxiliar na análise do material de pesquisa coletado a partir das entrevistas, pretendemos fazer uso de técnicas de análise de conteúdo. Assim, partimos de uma “leitura flutuante”, buscando identificar temas recorrentes no material que serviram de base para a criação de “categorias de análise”, com o objetivo de identificar, no material coletado, elementos relevantes que informam e compõem o uso da mineração de textos e dados na pesquisa, bem como a discussão e a implementação de políticas que interajam com essas tecnologias (Franco, 2005; Flick, 2009; Bardin, 2011).

A fim de não limitar o escopo de possíveis informações a serem encontradas no material selecionado, utilizou-se um sistema de definição de categorias *a posteriori*, a serem definidas conforme os padrões vão se relevando durante a pesquisa. Nesse mesmo sentido, tomou-se o “tema” como recorte para a categorização dos segmentos (ou “unidades de registro”) relevantes, e a seção do texto no qual determinada unidade de registro surge foi definida como a unidade de contexto. Tal como feito em trabalhos anteriores, visa-se seguir regras de exclusão mútua (uma unidade de registro por divisão); homogeneidade de registros e dimensões de análise; pertinência ao material e ao referencial teórico; objetividade e fidelidade (mesmas partes de um material sob a mesma divisão categorial devem ser codificadas da mesma maneira) (Bardin, 2011).

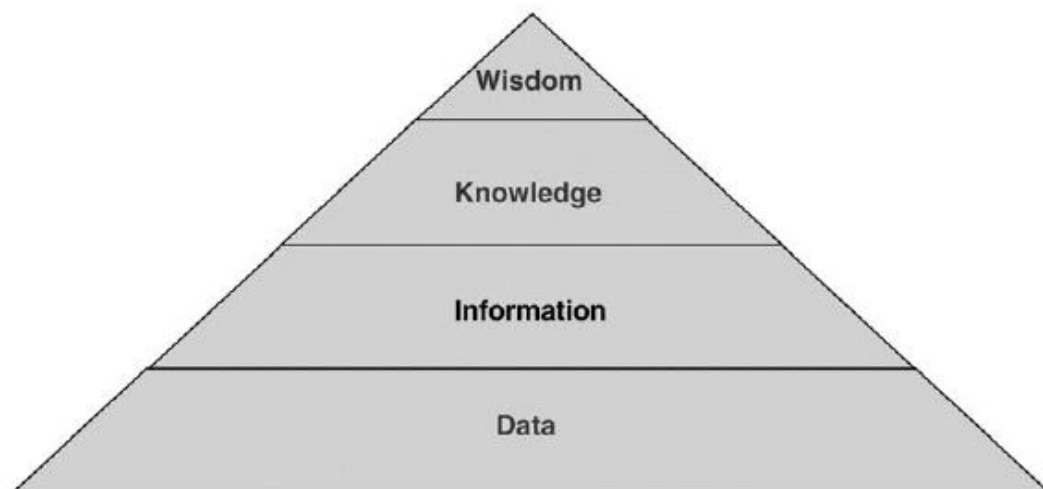
CAPÍTULO 1. MINERAÇÃO DE TEXTOS E DADOS E DIREITO AUTORAL

Esta tese trata, em seu âmago, da regulação sobre a posse e gestão de um objeto em especial: bancos de dados e seu conteúdo, quando estes são utilizados e gerados no ambiente de pesquisa. Portanto, convém explorar a natureza desse objeto e a origem do seu valor, como ponto de partida para o restante do trabalho. Para tanto, faz-se uso tanto do referencial teórico pertinente aos bancos de dados – exposto na seção de referencial teórico do capítulo anterior – como de informações obtidas a partir de observações documentais e empíricas realizadas para este trabalho.

1.1. A natureza dos dados e seus repositórios

Compreender a influência da propriedade intelectual sobre a mineração de textos e dados passa, necessariamente pelo entendimento dos objetos que esta tecnologia compõe e utiliza, a começar pelo menor componente desta estrutura: o dado. O seu significado de origem (do latim *datum*, ou “coisa dada”), remete à sua natureza: Rowley (2007, p. 166-168), em sua revisão bibliográfica, chegou a uma definição de dados enquanto representações das propriedades de objetos, eventos e seu ambiente – representações estas que existem de forma desconexa, de modo que não fazem sentido quando vistos de forma isolada. A geração de conhecimento depende, portanto, da agregação e contextualização desses dados de forma a transformá-los em informação, em um modelo que segue uma lógica crescente de complexidade, em uma estrutura chamada “pirâmide DYKW” (*data, information, knowledge and wisdom*, ou “dados, informação, conhecimento e sabedoria”) (Figura 3):

Figura 3: Pirâmide DIKW. Da base ao topo: “dados, informação, conhecimento e sabedoria”.



Fonte: Rowley, 2007.

Um exemplo prático de como isto funciona, a título de ilustração, ocorre quando descrevemos uma pessoa para alguém. Normalmente, cada indivíduo possui características que podem ser associadas a dados específicos (como nome, idade, altura, peso, ocupação, entre outros). Nenhum desses dados, por si só, permite a identificação precisa dessa pessoa – isto só se torna possível a partir do momento em que são considerados em conjunto.

A noção de dado como parte constitutiva da informação também constitui um entendimento que se reflete na própria conceituação de dados observada na legislação brasileira, a exemplo da Lei Geral de Proteção de Dados Pessoais (LGPD), que associa “dado” a “informação” (Brasil, 2018).⁹ Embora este trabalho parta do pressuposto de que ambos os termos são conceitualmente distintos, nota-se que a associação entre eles permanece, em função de questões técnicas a serem observadas na próxima seção.

Um subconjunto dos dados que merece destaque são os chamados metadados, ou “dados sobre dados”. São, literalmente, representações factuais acerca da estrutura e propriedades dos próprios dados ou do conjunto onde estão reunidos. Em um ambiente digital, estes metadados podem, por exemplo, indicar o formato de um arquivo, seu tamanho, um código que corresponda a uma cor específica, entre outros (Han, Pei e Kamber, 2011, p. 65; Kroenke et al, 2017, p. 19; Kelleher e Tierney, 2018, p. 9). No nosso exemplo acima, um metadado pode ser a própria observação de que “advogado” é uma profissão, que o espectro de alturas em consideração vai de 1 a 3 metros, e assim por diante.

⁹ “Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;” (Brasil, 2018). Aqui, associa-se o termo “pessoal” à relação de um certo dado com uma pessoa natural específica; por exclusão, resta a associação de “dado” e “informação”, que aqui se confundem.

Outro subconjunto de interesse para este trabalho é a visão de dados científicos, definidos pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) como

Registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para a pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar resultados de pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do sujeito ora investigado (OCDE, 2007, p. 13. Tradução e destaque nossos).¹⁰

As definições acima parecem levar à percepção de que os dados são, então, meras observações factuais e objetivas de artefatos. É uma noção ainda presente na atualidade, a exemplo de Kelleher e Tierney (2018, p. 39; 48-49) quando estes definem dados como abstrações feitas a respeito de uma entidade real, sob a forma de valores ou atributos – os quais normalmente existem de forma não-estruturada, como conjuntos de instâncias, objetos ou indivíduos em diferentes formatos e naturezas, necessitando passar por um trabalho de organização para se tornarem estruturados, seguindo um formato em comum que facilite o armazenamento, organização, busca e manipulação desses dados.

Entretanto, tanto a definição dos autores como a definição dada pela OCDE já sugerem que essas observações “abstratas” podem se materializar de diversas maneiras – na verdade, para Gitelman (2013, p. 6), dados *necessitam* de uma manifestação material, normalmente gráfica, para existirem e serem mobilizáveis. Borgman (2011, p. 6) afirma que a definição de “dados” não é tão simples quanto parece, justamente por conta desta variedade de formas: a autora aponta, inclusive, para documentos de pesquisa que adotam uma interpretação extensiva do que conta como “dado”, o que incluiria não apenas textos, imagens e áudio, mas também elementos como jogos e simulações.

Ao fim e ao cabo, isto também implica em dizer, tal como sugere a última frase na definição da OCDE, que o fato de um dado representar fatos não implica em dizer que esta observação seja neutra ou impessoal. Na verdade, é possível inverter esta lógica e constatar que própria consideração de que um determinado material constitui um tipo de dado já depende, por si só, do observador – sendo esta nomeação, em si mesma, “um ato acadêmico” Borgman (2011, p. 5-6).¹¹ De forma similar, Gitelman (2013, p. 3) afirma que, diferentemente de eventos, dados

¹⁰ “(...) factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated (...).” (OCDE, 2007, p. 13).

¹¹ “Data may exist only in the eye of the beholder: the recognition that an observation, artifact, or record constitutes data is itself a scholarly act. Data curators, librarians, archivists, and others involved in data management may be offered a collection that is deemed data by the collector, but not perceived as such by the recipients. Conversely, an investigator may be holding collections of materials without realizing how valuable they may be as data.” (BORGMAN, 2011, p. 5).

são gerados a partir de sistemas de conhecimento, de forma que disciplinas diferentes terão formas diferentes de conceber dados.¹² Esta informação é importante quando consideramos as visões sobre uso de dados na academia, a serem abordadas mais adiante.

Ainda neste sentido, Gitelman (*ibid.*, p. 6-12) ainda aponta que, além de os dados serem abstratos (necessitando, portanto,) e agregativos (existirem em fragmentos a serem agrupados), estes também possuem uma função retórica que os distingue dos fatos aos quais se relacionam. Enquanto fatos em si operam em uma esfera ontológica (são verdadeiros ou falsos, existem ou não existem), dados surgem como elementos argumentativos para um posicionamento: são bons ou ruins, completos ou incompletos. Dados diferentes podem ser usados para se chegar a conclusões diferentes.

Isto também pode ser observado a partir das conclusões de Rosenberg (2013): o autor observa que, ao longo de sua análise histórica sobre o uso do termo “dado” (mais especificamente, “data”), esta função retórica surge de diferentes maneiras ao longo do tempo: inicialmente premissas fornecidas antes da argumentação, embora, a partir do séc. XVIII, tenham começado a ser entendidos não como premissa, mas como fatos observados a partir de uma investigação. Assim, a própria noção de dados é algo mutável, assumindo concepções diferentes ao longo do tempo.

As conceituações acima levam à conclusão de que dados são materiais essenciais para a construção de conhecimento, mas eles não são capazes de fazê-lo sozinhos: existem de forma simbiótica com métodos de conceituação, agregação, organização e mobilização que insiram estes fragmentos abstratos em um conjunto coerente e concreto. É neste contexto que se inserem as ferramentas de coleta e análise de dados, dentre as quais destacamos a mineração de textos e dados.

1.2. Mineração de textos e dados

O uso de técnicas sistemáticas de coleta e análise de dados como um elemento importante para a construção de resultados relevantes não é um fenômeno recente. No século 18, o filósofo e teólogo Joseph Priestley já fazia uso extensivo de constelações quantitativas e qualitativas de “dados” e conjuntos de dados relacionados a dados para fundamentar suas pesquisas (Rosenberg, 2013). Da mesma forma, Garvey (2013) apresenta o livro “American Slavery As It Is: Testimony of a Thousand Witnesses” como um exemplo de produto, apenas um século

¹² “Events produce and are produced by a sense of history, while data produce and are produced by the operations of knowledge production more broadly. Every discipline and disciplinary institution has its own norms and standards for the imagination of data (...).” (Gitelman, 2013, p. 3).

após o trabalho de Priestley, de um processo sistemático e metucioso de coleta, organização, correlação e apresentação de dados, seguindo um propósito e uma narrativa específicos – ao qual a própria autora se refere como “mineração de dados”.

De forma similar, a dificuldade de se coletar e analisar grandes quantidades de dados também não é novidade: Dean (2014, p. 8-9, 11-14) e Martens (2018, p. 6-7) apontam que, antigamente, por mais importante que a análise de dados fosse, não se dispunha de ferramentas capazes de observar e contextualizar estes dados de forma prática e factível, de forma que era comumente necessário empregar metodologias que permitiam alcançar resultados médios ou “por aproximação”, como na análise por amostragem.

Isto foi potencializado por um processo de crescimento exponencial na capacidade de processamento e de armazenamento de dados, bem como aumento de largura da banda, ocorrendo de forma mais acentuada a partir das últimas décadas do século XX. Tal expansão permitiu que a geração e transmissão de dados se desse em quantidade, facilidade e amplitude cada vez maior. Isto nos leva a um universo de dados tão vasto que não pode ser adequadamente explorado com o uso de tecnologias de armazenamento e processamento convencionais (Tigre e Pinheiro, 2019).¹³ Tem-se, então, o conceito de *megadados*, mais comumente conhecidos como *Big Data*, que são “conjuntos de dados que não podem ser percebidos, adquiridos, geridos e processados por ferramentas tradicionais de *software* e *hardware* em um tempo tolerável” (Chen, Mao e Liu, 2014, p. 173).

O levantamento realizado por Antoniutti (2015, p. 67) aponta para diversas divergências na literatura acerca de quais seriam as características principais dos *megadados*, mas cinco elementos recorrentes podem ser observados: o primeiro é o seu grande volume, normalmente vários objetos reunidos ao longo de um tempo que pode compreender anos. Também existem em grande variedade, sendo possível encontrar diversos formatos de arquivos dentro de um mesmo conjunto. Também são produzidos em grande velocidade, quase instantânea,¹⁴ e se

¹³ Segundo os autores, estes elementos compõem um tripé responsável pelo desenvolvimento de uma economia intensiva em informação nas últimas décadas. A primeira perna deste tripé, o aumento do processamento, estaria relacionada à "Lei de Moore", a qual prevê que o poder de processamento dos chips dobraria a cada 18 meses (embora a sustentação desse ritmo pudesse mudar no longo prazo). A segunda perna, a capacidade de armazenamento, diz respeito ao desenvolvimento de novos chips de memória e ao desenvolvimento da computação em nuvem, que permitiram o surgimento de *data centers* capazes de comportar uma demanda cada vez maior por dados. Finalmente, o terceiro tripé é a largura da banda digital, ou o aumento da capacidade de transmissão de dados, proporcionada por avanços de infraestrutura (ex.: fibras ópticas e comunicações sem fio) (Pinheiro e Tigre, 2019).

¹⁴ A título de comparação, estima-se que, no ano de 2011, em apenas dois dias eram gerados cerca de 1,8 zetabytes (equivalentes a aproximadamente 1,8 milhão de terabytes, ou 1,8 trilhão de gigabytes), mais do que o total de dados gerados do início da civilização humana até o ano de 2003 (CHEN, 2014, p. 172; Dean, 2014, p. 6).

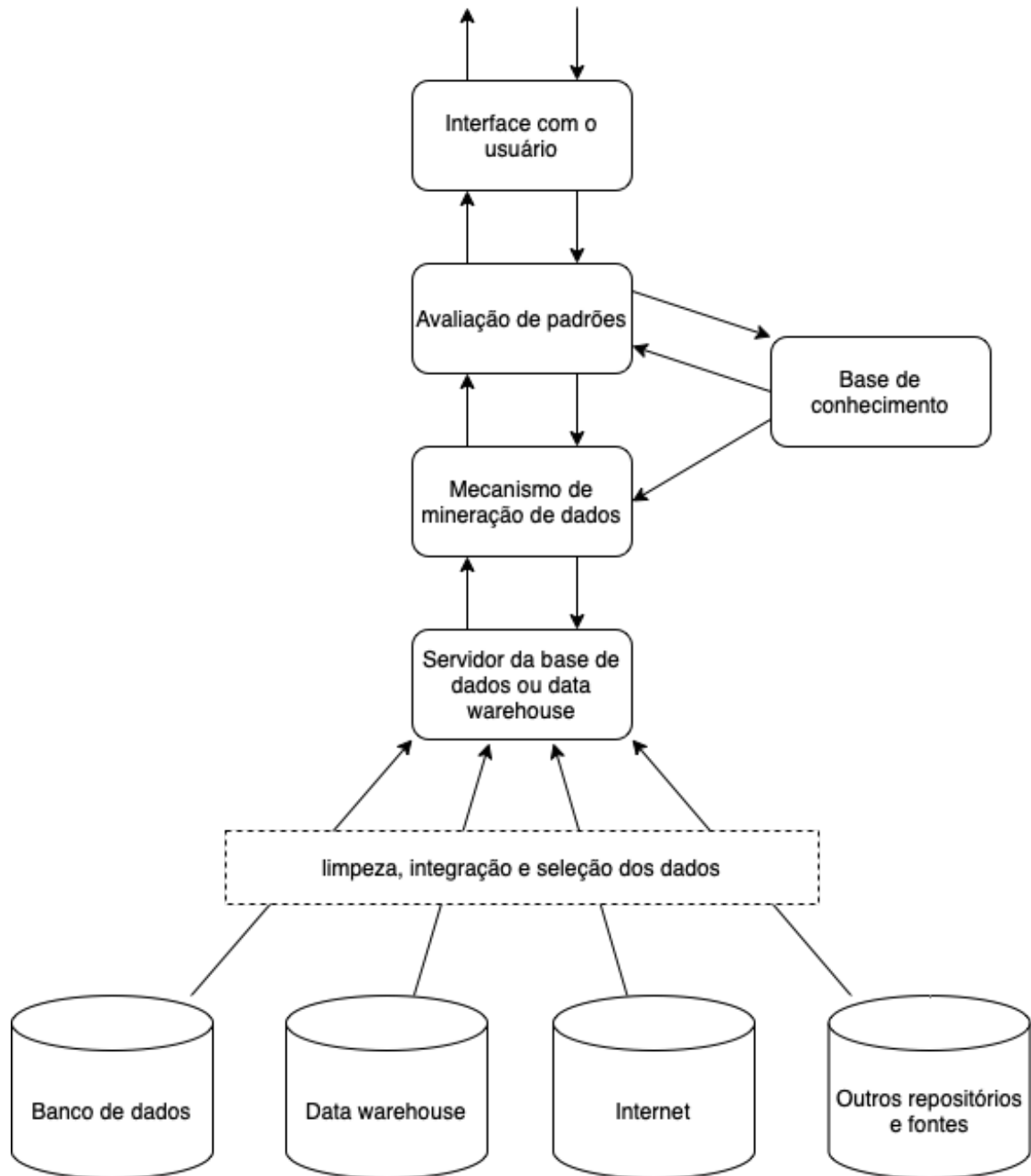
defasam com rapidez, demandando um processamento cada vez mais veloz (Dean, 2014, p. 4). Outros elementos incluídos em alguns conceitos de *Big Data* são a veracidade, ou a importância de se confirmar a qualidade, a autenticidade e a reputação dos dados coletados, (Gervais, 2019, p. 23), e o valor – quanto maior o conjunto de dados, e quanto melhor é a capacidade de analisá-los de forma completa, eficiente e detalhada, mais valor eles possuem.

É neste contexto que ganha destaque a mineração de textos e dados, ou MTD (em inglês, *text and data mining*, ou TDM) como a conhecemos hoje. Também conhecida como descoberta de conhecimento a partir de dados, ela é essencialmente definida como um processo pelo qual padrões e correlações em grandes volumes de dados são pesquisados para obter diversas informações (Han, Pei e Kamber, 2011; Kroenke et al, 2016; Kelleher e Tierney, 2018).

Mais especificamente, Han, Pei e Kamber (2011, p. 6-7) a descrevem como uma parte específica de um processo composto por diversas etapas – as quais são aqui expostas de forma sequencial, mas que, na prática, podem ocorrer até mesmo simultaneamente. São elas:

- os dados são coletados de várias fontes e "limpos" (dados desnecessários, corrompidos ou redundantes são removidos da coleção);
- os dados são integrados em um único local para facilitar a análise;
- os dados relevantes são selecionados do banco de dados utilizado pelo analista e
- transformados em um formato mais adequado para análise (por exemplo, digitalização de uma pintura ou texto de um livro, convertendo o material em um formato digital legível por máquina);
- a máquina ou a pessoa responsável pela análise começa a examinar os dados em busca de padrões e correlações que possam surgir (esta é a mineração de dados propriamente dita) e
- faz medições para identificar padrões que possam ser de interesse;
- apresentação dos dados ao usuário final e integração destes ao conjunto de dados existentes para processamento posterior, gerando um *loop* contínuo de análise.

Figura 4: Arquitetura de um típico processo de mineração de textos e dados.



Fonte: Adaptado de Han, Pei e Kamber (2011).

Um termo relacionado e baseado nos mesmos princípios é a *mineração de textos*, definida como um conjunto de técnicas que buscam extrair informações úteis de documentos escritos (como livros, páginas da Web, e-mails, relatórios etc.) por meio da identificação de padrões de destaque que emergem dos dados não estruturados existentes nesses textos (Truyens e van Eecke, 2014). Como os dados podem ser interpretados de forma a incluir conteúdo textual e tanto a mineração de texto quanto a mineração de dados são funcionalmente iguais, compartilhando métodos e operações, esses termos são frequentemente citados como um só: *text and*

data mining (TDM)¹⁵ ou, como este trabalho adotará como padrão, *mineração de textos e dados* (MTD).

Outro termo relacionado é a raspagem (*scraping*, em inglês) de dados, que se refere especificamente a técnicas de busca e extração automatizada de dados, através de programas ou robôs (ou *bots*). Está mais diretamente ligada ao processo de coleta de dados (Murthy et al, 2013; Ashraf, 2017), embora seja possível ver o termo sendo usado de forma sinônima à mineração em si – como visto, por exemplo, em Henrys (2021) e em diversos termos de uso expostos ao longo deste trabalho.

Enquanto parte essencial do processo de análise de dados, sua presença está, consequentemente, associada à maioria das atividades que fazem uso intensivo de dados atualmente: desde estabelecer caminhos inexplorados na pesquisa científica e prever futuras descobertas, além de auxiliar na gestão de governos e empresas, no desenvolvimento de novas tecnologias, no aprimoramento de sistemas de segurança e, na área da saúde, introduzindo melhores diagnósticos médicos e fazendo uso de técnicas de sequenciamento genético para entender melhor várias doenças (Chen, Chiang e Storey, 2012, p. 1168-1172; McAfee e Brynjolfsson, 2012, p. 6-8).

Quanto a este último ponto, a mineração de textos e dados desempenhou especial papel no combate à pandemia de COVID-19. Souza, Schirru e Alvarenga (2020) destacaram o uso da coleta e análise de dados em diversos “fronts”. Um deles foi a construção de observatórios de monitoramento da doença, centralizando informações de diversas fontes que observavam a evolução da doença. Um desses observatórios foi estabelecido pela Universidade Johns Hopkins (2020), utilizando dados governamentais e de hospitais para apontar números de casos e mortes, variações nos números ao longo do tempo, entre outros.

Outra frente científica foi voltada para a compreensão do próprio vírus. Um exemplo desse esforço pôde ser visto na plataforma NextStrain (2024), que se dedica a cruzar e interpretar dados genéticos obtidos de pesquisas obtidas em repositórios abertos ou aos quais os membros da plataforma possuem acesso. O resultado final é a construção de um mapa interativo sobre a evolução genética e da propagação de populações de vírus, bactérias e outros patógenos, a fim de auxiliar virologistas, epidemiologistas, funcionários de saúde pública e cientistas no entendimento de epidemias e na implementação de respostas às mesmas.

¹⁵ Por exemplo, ver Martens (2018), Geiger e Izyumenko (2019) e Sag (2019).

Mais um exemplo de destaque nas pesquisas genéticas sobre a COVID-19 pôde ser visto quando os primeiros casos da doença chegaram ao Brasil: uma equipe de pesquisadores do Instituto Adolfo Lutz e da Universidade de São Paulo foi capaz de obter o genoma do vírus contido nas amostras de alguns dos pacientes em apenas 24 horas – tudo isto combinando registros médicos com informações genéticas compartilhadas através de uma plataforma de divulgação de pesquisas médicas, a GISAID (Jesus et al, 2020).¹⁶ Aqui, nota-se que tais pesquisas possuem desdobramentos práticos fundamentais: de forma mais contundente, estas descobertas possibilitam identificar padrões nas variantes de patógenos e como estes mudam ao longo do tempo – o que é um passo crucial para o desenvolvimento de tratamentos e vacinas (Brasil, 2020a).

Para além das pesquisas médicas, Flynn e Vyas (2022) fornecem vários outros exemplos: ajudar os acadêmicos a encontrar, ler e analisar materiais bibliográficos para revisões de literatura, desenvolver ferramentas de tradução em diferentes idiomas, usar a análise linguística para avaliar as diferenças nas representações de indivíduos de gênero na literatura, entre outros.

Aqui, contudo, destaca-se que o campo da ciência não o único envolvido com atividades de pesquisa intensiva em dados. No campo da preservação, já se reconhece o papel das tecnologias de coleta e análise de dados, seja como parte do processo de conservação em si (por exemplo, mediante processos de digitalização), seja como ferramenta de aprimoramento da gestão e acesso às coleções – o que se intensificou nos últimos anos, com a COVID-19 (Noehrer et al, 2021; Marques, J., 2022). Isto se torna especialmente relevante em se tratando de projetos com arquivos nativamente digitais, a exemplo do Wayback Machine: uma ferramenta que realiza uma raspagem periódica de todos os websites publicamente disponíveis na internet, arquiva-os e disponibiliza-os para visualização, permitindo o acesso a certos conteúdos e documentos mesmo após terem sido alterados ou suas páginas saírem do ar. O projeto é mantido pelo Internet Archive (2024), entidade sem fins lucrativos dedicada à conservação digital de arquivos.

A mineração de textos e dados também tem se mostrado uma ferramenta útil para outro tipo de pesquisa: a jornalística. Um exemplo disso está no crescimento do chamado “jornalismo

¹⁶ O GISAID é um exemplo interessante de plataforma para o compartilhamento de dados científicos: embora se posicione como uma alternativa às plataformas de "domínio público", as quais alega não serem capazes de garantir a devida proteção ao trabalho de pesquisadores (ver: <https://www.gisaid.org/about-us/history/>), seus termos de uso dão acesso livre de *royalties* aos dados compartilhados em seu banco de dados, ao mesmo tempo em que cria regras simples para reconhecer os criadores originais dos dados e restringir seu uso por terceiros. Para obter mais informações, ver: <https://www.gisaid.org/registration/terms-of-use/>.

de dados”, que utiliza técnicas computacionais de coleta e análise para contar histórias com o apoio de dados: seja revelando informações ali ocultas, usando-os enquanto ferramenta investigativa, como objetos de escrutínio por si sós, ou simplesmente para dar maior robustez ao que é reportado (Coddington, 2015; Rogers, Schwabish, and Bowers, 2017; Heravi, 2019; Bounegru e Gray, 2021).¹⁷

Outra aplicação na mesma área está na checagem de fatos e verificação de notícias falsas – a exemplo da agência brasileira Aos Fatos, que possui um Radar baseado em algoritmos com curadoria constante de linguistas. O programa coleta publicações e posts em diversos sites, buscando palavras-chave que correspondam a conteúdos tipicamente associados a informações falsas sobre diversos temas (Aos Fatos, 2023). A partir daí, segundo anteriormente indicado em sua metodologia, a agência verificaria uma determinada notícia, comparando-a com as informações de seu repositório, e atribuiria a este uma nota de 1 (incorreta ou não-confiável) a 10 (extremamente correta ou confiável) (Aos Fatos, 2021).

Isto posto, uma aplicação em especial vem recebendo destaque pronunciado nos últimos anos: o uso da mineração de textos e dados para o treinamento de sistemas de inteligência artificial (IA), que, resumidamente, são definidos pela OCDE da seguinte forma:

Um sistema de IA é um sistema baseado em máquinas que, para objetivos explícitos ou implícitos, infere, a partir das informações que recebe, como gerar resultados tais como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após sua implementação. (OCDE, 2024, p. 4. Tradução feita pelo autor).

De forma mais técnica, o mesmo memorando (OCDE, 2024, p. 8) situa a vasta gama de modelos IA em duas modalidades: a primeira é o *aprendizado de máquina*, ou *machine learning* (ML), focado em permitir a uma máquina ou algoritmo extrair informação dos dados e usá-la como ponto de partida para realizar tarefas e atingir objetivos de forma flexível (Samuel, 1959, p. 1; Kaplan e Haenlein, 2018, p. 7). A Universidade de Berkeley descreve o ML em três partes: um processo de decisão propriamente dito, baseado em cálculos ou outras etapas que farão a identificação de padrões; uma função de erro que avalia o resultado em comparação com exemplos estabelecidos e, por fim, um sistema de otimização, que atualiza o processo de

¹⁷ Para maiores exemplos de aplicações do jornalismo de dados e os aspectos técnicos associados, ver Bounegru e Gray, 2021.

decisão para reduzir o erro (UC Berkeley, 2022).¹⁸19 A segunda modalidade apontada pela OCDE (2024, p. 8) são os modelos simbólicos, que utilizam representações de variáveis lógicas ou probabilísticas para inferir seus resultados – por exemplo, um sistema focado em soluções para indústria pode fazer uso de variáveis relacionadas a maquinário, mão-de-obra, fábricas e produtos. Tais representações podem, inclusive, ser construídas com auxílio de *machine learning*, formando modelos mistos.²⁰

Russel e Norvig apontam que os avanços no desenvolvimento de sistemas de inteligência artificial estariam diretamente relacionados ao desenvolvimento da mineração de textos e dados no âmbito digital – e, por conseguinte, a IA é um dos campos que mais avançaram enormemente nas últimas décadas justamente com o aprimoramento das técnicas de mineração de textos e dados e com a expansão da capacidade de armazenamento de informação. Com efeito, para os autores, os dados ocupariam um papel muito mais importante no desenvolvimento de inteligência artificial do que a própria escolha do algoritmo (Russel e Norvig, 2013, p. 26-27).²¹2223 Atualmente, diversas tecnologias de ponta, como veículos autônomos,

¹⁸ “The basic concept of machine learning in data science involves using statistical learning and optimization methods that let computers analyze datasets and identify patterns (...). Machine learning techniques leverage data mining to identify historic trends and inform future models.

The typical supervised machine learning algorithm consists of roughly three components:

1. A decision process: A recipe of calculations or other steps that takes in the data and “guesses” what kind of pattern your algorithm is looking to find.
2. An error function: A method of measuring how good the guess was by comparing it to known examples (when they are available). Did the decision process get it right? If not, how do you quantify “how bad” the miss was?
3. An updating or optimization process: A method in which the algorithm looks at the miss and then updates how the decision process comes to the final decision, so next time the miss won’t be as great.” (UC Berkeley, 2022).

¹⁹ Subtipos desse sistema são as *redes neurais*, que buscam imitar o padrão de neurônios do cérebro humano em processos decisórios (Haykin, 2001); o *aprendizado profundo*, ou *deep learning*, que busca reproduzir o modelo aprender padrões mais complexos e extrapolá-los, sem a necessidade de prévia catalogação e organização (Najafabadi et al, 2015); as *redes artificiais generativas*, ou *generative artificial networks* (GANs), usadas na geração de conteúdos (como imagens) e a verificação de sua veracidade como forma de treinamento (Drexl et al, 2019), e o *processamento de linguagem natural*, que visa decodificar padrões linguísticos e serve de base para os tradutores on-line atualmente disponíveis. Para mais informações, ver Schirru (2020).

²⁰ Há, ainda, diversas outras modalidades apontadas na literatura, como os algoritmos genéticos, onde diferentes soluções para um problema competem entre si a fim obter resultados melhores a cada nova iteração (Mitchell, 2009) e os sistemas especialistas, que buscam codificar a expertise de determinados especialistas em uma área para a resolução de problemas na mesma (Luger, 2004).

²¹ “Much of the work on neural net in the 1980s was done in an attempt to scope out what could be done and to learn how neural nets differ from “traditional” techniques. Using improved methodology and theoretical frameworks, the field arrived at an understanding in which neural nets can now be compared with corresponding techniques from statistics, pattern recognition, and machine learning, and the most promising technique can be applied to each application. As a result of these development,s so-called data mining technology has spawned a vigorous new industry” (Russel e Norvig, 2013, p. 26).

²² “Throughout the 60-year history of computer science, the emphasis has been on the *algorithm* as the main subject of the study. But some recent work on AI suggests that for many problems, it mkes more sense to worry about the *data* and be less picky about what algorithm to apply. This is true because of the increasing availability of very large data sources (...)” (*Ibid.*, p. 27).

²³ Aqui, cabe um parêntesis para destacar que isto não descarta a importância de se considerar a arquitetura dos algoritmos, especialmente considerando que a inteligência artificial é, ainda, desenvolvida por pessoas, seguindo

reconhecimento de fala, planejamento de cronogramas, agendas e logística, robótica e tradução por máquina (*machine translation*, ou MTL), fazem uso de sistemas de inteligência artificial (*Ibid.*, p. 28-29).

Para que a realização de toda esta gama de atividades seja possível, o processo de coleta e mineração de textos e dados requer uma complexa arquitetura, da qual apenas a camada mais externa – a interface com o usuário, pela qual se interage com o sistema e obtém-se as informações que são apresentadas ao final da análise – é diretamente visível para um observador leigo. Há também que se levar em conta os servidores responsáveis pela busca dos dados relevantes, a base de conhecimento que informa os parâmetros a serem utilizados no processamento, busca e avaliação de padrões, o mecanismo de mineração em si (que contém módulos responsáveis por realizar as correlações e previsões), o módulo de avaliação dos padrões encontrados que será responsável pela seleção dos resultados relevantes e, finalmente, a fonte da qual os dados se originam. Esta estrutura implica na necessidade da construção de repositórios que servirão para armazenar os dados coletados, ao mesmo tempo em que, por muitas vezes, também podem servir como fontes para processos de coleta por terceiros. Tais estruturas são os *bancos de dados*, ou *bases de dados* (Han, Pei e Kamber, 2011, p. 7-9), ou ainda *datasets*, elemento de especial interesse para este trabalho.

De forma mais específica, Kroenke et al (2018, p. 3-12) definem bancos (ou bases) de dados como estruturas criadas a partir de informações coletadas de várias fontes, cuja função é permitir a preservação, acesso e cruzamento mais eficientes dos dados coletados. Essa organização inclui a criação e o armazenamento de metadados, índices e descrições dos aplicativos usados. Esses agrupamentos ou coleções de dados e metadados fazem parte de um sistema maior que inclui programas de computador dedicados à criação, ao processamento e à administração desses bancos de dados - que compõem o sistema de gerenciamento de banco de dados (DBMS) -, os aplicativos que servem de interface entre os usuários e o DBMS e os próprios usuários.

Bancos de dados podem existir em formato relacional (ou SQL), com tabelas que representam relacionamentos entre atributos e entidades através de linhas e colunas (Han, Pei e Kamber, 2011, p. 10-12; Kroenke et al, 2018; Kelleher e Tierney, 2018, p. 7-8), ou, como é cada vez mais comum, em formato não-relacional (NoSQL), em que os dados são simplesmente

parâmetros e escolhas permeadas de personalidade e de elementos sociais, políticos e econômicos – assim como o próprio processo de geração de dados, conforme discutido anteriormente. Para um exemplo de como estas questões são construídas por – e, em troca, constroem – esses elementos, ver o trabalho de Lobato (2022), sobre o papel dos algoritmos na construção da (in)segurança no Sul Global.

representados como objetos separados com atributos próprios, sendo “etiquetados” com termos que permitam a busca e a correlação com outros dados sem a necessidade de ajustá-los a nenhum formato fixo de linhas e colunas – o que é útil quando se trabalha com arquivos que se apresentam em uma grande variedade de formatos (vídeo, áudio, texto etc.) (Kelleher e Tierney, 2018, p. 9-10). Essa distinção será importante ao tratarmos da transparência de sistemas de IA mais adiante.

Finalmente, bases de dados ocupam papel de destaque na mineração de textos e dados, uma vez que podem aparecer duas vezes na arquitetura do processo de coleta e análise, com propósitos distintos: são criados para armazenamento, consulta e outras atividades operacionais, muitas vezes servindo como ponto de origem para a obtenção de material para análise. Já durante o processo de limpeza, integração e transformação, estes elementos são comumente reunidos dentro dos armazéns de dados, ou *data warehouses*, que são bases de dados criadas com o propósito específico de reunir o conteúdo de outras fontes em um único *corpus* de pesquisa, de forma a facilitar a modelagem e análise dos dados, além de servir como ponto de armazenamento das informações resultantes do processo de mineração de dados, permitindo o reuso do material em novas análises (Han, Pei e Kamber, 2011, p. 105-107; Kelleher e Tierney, 2018, p. 8-9; Kroenke et al, 2018, p. 492-494).

Dado o impacto crescente das tecnologias de mineração de textos e dados, parece lógico concluir, portanto, que a posse e uso de grandes bases de dados seja atualmente vista como ativo essencial para seus detentores. Não surpreende, portanto, a consolidação de uma realidade econômica marcada pela busca em converter cada vez mais a vida em dados exploráveis comercialmente (Couldry e Mejias, 2019a; 2019b) e pelo chamado *capitalismo de plataforma* (Srnicsek, 2017), definido pelo crescente poder econômico de detentores de grandes empresas de tecnologia que se colocam como intermediários responsáveis por formatar e mediar mercados e relações sociais – usando-se do conteúdo gerado por seus usuários para gerar cada vez mais dados e utilizá-los, por exemplo, como moeda de troca para entidades interessadas em acessar este tipo de informação

Este cenário, somado ao fato de que a construção e manutenção de bases de dados costuma requerer investimentos consideráveis de tempo, recursos técnicos e, mais importante, capital financeiro e humano (Dean, 2014, p. 12), motiva crescentes esforços empreendidos por donos de grandes repositórios no sentido de controlar o acesso e uso de suas bases por terceiros (Stucke e Grunes, 2015, p. 3). Pode-se entender este receio como baseado na ideia de que o conteúdo e a estrutura de bases de dados, por sua natureza abstrata, podem ser copiadas sem

dificuldade – de fato, reproduzir uma base de dados inteira representaria um custo muito menor do que criar um repositório do zero. Logo, os autores percebem a presença de fortes incentivos para a implementação de iniciativas de “fechamento” ou “cerceamento” desta informação valiosa.

Apesar de opiniões divergentes no sentido de que a natureza efêmera e replicável dos dados dificultaria este processo de isolamento (Sokol e Comerford, 2016), a preocupação de que este cenário pudesse levar a um subinvestimento na criação e manutenção de bancos de dados serviu de justificativa para a adoção de medidas institucionais, em nível nacional e global, com o objetivo de mitigar essa questão (Wu, 2008, p. 35). Um desses mecanismos, como será visto ao longo deste trabalho, é a criação de direitos de propriedade intelectual; mais especificamente, pela via do Direito Autoral.

1.3. A proteção de bases de dados e seu conteúdo pelo Direito Autoral

Na década de 1990, o Direito Autoral começou a tratar os dados e seus repositórios como elementos cujas características e valor econômico os qualificariam a alguma tipo de resguardo legal. Assim, desde o Acordo TRIPS (Organização Mundial do Comércio, 1996), as bases de dados fazem parte do grupo de obras protegidas por Direito Autoral.²⁴

O Brasil seguiu este movimento, tendo sido esta uma das novidades introduzidas na Lei n. 9.610/98. Trata-se de uma atualização em relação a um tipo de obra mais antiga: as compilações de obras, que já recebiam proteção na lei n. 5.988/73, a versão anterior da Lei de Direitos Autorais.²⁵ O texto atual foi para além das simples coletâneas de obras, incluindo, tal como em TRIPS, as bases de dados como material protegido.²⁶ Contudo, nota-se que, tanto em TRIPS como no Brasil, este direito não se estende ao conteúdo da base de dados. Isto quer dizer

²⁴ “Article 10
Computer Programs and Compilations of Data
(...)”

2. Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself” (ORGANIZAÇÃO MUNDIAL DO COMÉRCIO, 1994, p. 324).

²⁵ “Art. 7º Protegem-se como obras intelectuais independentes, sem prejuízo dos direitos dos autores das partes que as constituem, as coletâneas ou as compilações, como seletas, compêndios, antologias, enciclopédias, dicionários, jornais, revistas, coletâneas de textos legais, de despachos, de decisões ou de pareceres administrativos, parlamentares ou judiciais, desde que, pelos critérios de seleção e organização, constituam criação intelectual” (Brasil, 1973).

²⁶ “Art. 7º São obras intelectuais protegidas as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro, tais como: (...) XIII - as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual (...)”

§ 2º A proteção concedida no inciso XIII não abarca os dados ou materiais em si mesmos e se entende sem prejuízo de quaisquer direitos autorais que subsistam a respeito dos dados ou materiais contidos nas obras” (Brasil, 1998b).

que a proteção dada às compilações não dá ao seu dono o status de detentor de direitos autorais sobre qualquer obra protegida ali inserida – os quais ainda pertencem a seus respectivos titulares. Pela mesma lógica, quaisquer materiais em domínio público existentes em uma base de dados assim permanecem, devendo ter seu acesso e uso deixados livres pelas regras do Direito Autoral.

Outro ponto de destaque do texto legal é que apenas se protegem as bases de dados que sejam consideradas obras intelectuais em razão da seleção, organização ou disposição do seu conteúdo. Mas o que qualifica um material como “obra intelectual”?

A LDA em si não é clara com relação a isto, simplesmente considerando obras protegidas todas as chamadas “criações do espírito”: obras artísticas, literárias e científicas, que possuem, por si sós, valor estético ou enquanto conhecimento - diferente de obras “utilitárias”, cujo valor é intrinsecamente associado à realização de uma função (Bittar, 2008, p. 21-22). Entretanto, um elemento central para esta definição parece ser a noção de originalidade, a qual será discutida a seguir.

Com efeito, o critério de originalidade no Direito Autoral é mutante, aplicado de forma distinta no espaço e no tempo, mas que, no geral, adquire dois sentidos principais. O primeiro, segundo Barbosa, D. (2005), é a mera novidade, que não é absoluta – já que toda obra é inspirada em elementos técnicos ou culturais anteriores – mas que implica, simplesmente, em ser concebida de forma diversa do que já existe. Para Deveci (2004, p. 180-181), esta noção de “novidade” já seria distinta da “originalidade” considerada no Direito Autoral: aquela constituiria uma relação entre uma criação (invenção) e a informação disponível em domínio público, enquanto a última expressaria a existência de uma relação entre criação e criador (autor).

Tomando como base o conceito anterior, considera-se que um texto, seja ele de cunho artístico ou científico, revela diretamente a adoção de escolhas que refletem, de uma forma ou de outra, a personalidade do seu autor. Revelam, portanto, uma questão subjetiva e, por esta definição, um elemento de originalidade. Este elemento subjetivo é inclusive sugerido no texto da Diretiva 2006/116/CE do Parlamento Europeu, o qual, em seu Recital 16, a respeito das obras fotográficas, aponta que estas devem, no sentido expresso na Convenção de Berna, ser consideradas originais sempre que constituírem uma criação intelectual do autor, refletindo a sua personalidade, independentemente de mérito ou finalidade.²⁷

²⁷ “A protecção das fotografias nos Estados-Membros é objecto de regimes diferentes. Uma obra fotográfica, na acepção da Convenção de Berna, deve ser considerada original sempre que for criação intelectual própria do respectivo autor, reflectindo a sua personalidade, sem que outros critérios, tais como o mérito ou a finalidade, sejam

Finalmente, há um outro conceito de originalidade apresentado por Barbosa, D. (2005, p. 4-6), que possui, por sua vez, um caráter mais objetivo – o que é de especial importância quando analisamos obras de cunho mais utilitário, como softwares e bancos de dados digitais. Aqui, uma obra se qualificaria como original em função de sua “distinguibilidade”: não pode ser banal ou comum, devendo trazer algo novo por si só de forma a ser inconfundível com outras obras do mesmo gênero.

Barbosa, P. (2018) dá continuidade a essa discussão, destacando outro ponto essencial para a compreensão do conceito de originalidade dentro do Direito Autoral – como observá-la tendo em mente não apenas o próprio processo de criação cultural, no qual o novo se apoia sobre o acervo cultural que o antecedeu, como também o fenômeno de massificação cultural, no qual obras repetem cada vez mais tropos, clichês e lugares-comuns que não são específicos a uma produção, mas a gêneros ou mídias inteiras? Nesse sentido, o autor traz a noção de “contributo mínimo”, destacando o “original” como aquilo que distingue uma obra das demais do mesmo gênero. Assim, conclui-se que, a bem do interesse público, a originalidade necessária para atrair a proteção do Direito Autoral não se aplica sobre uma obra como um todo, mas apenas sobre os elementos específicos da mesma que a diferenciam dos seus “pares”.²⁸

Isto carrega similaridades com o conceito de originalidade no Direito Autoral dos Estados Unidos, que tem duas características principais. A primeira é que o material deve ter sido fruto de uma criação independente: em outras palavras, não constituir mera cópia, tendo resultado de uma produção intelectual autônoma do seu criador. O segundo fator é o chamado “mínimo grau de criatividade”. O que constitui este mínimo grau se transformou ao longo do tempo, embora seus contornos atuais tenham sido limitados pelo caso *Feist Publications, Inc. v. Rural Tel. Service Co., Inc.* (conhecido simplesmente como o “caso Feist”), o qual lidou, coincidentemente, com a originalidade de bases de dados.²⁹

tomados em consideração. A proteção das outras fotografias deve poder ser regulada pela legislação nacional” (União Europeia, 2006).

²⁸ “Desse breve incurso filosófico também se pode concluir: (a) que toda criação advém de uma experiência que lhe antecede, (b) que a constatação da originalidade não significa uma ruptura com o que já exista e (c) que também o conceito de reprodução/plágio⁵² deve levar em conta o grau de distância entre todo o acervo pretérito e o salto qualitativo do contributo mínimo da obra, supostamente copiada. A última assertiva é deveras importante, pois a se levar rigidamente o contexto de tutela a obra anterior, dificilmente alguma criação estaria impune do entendimento sobre reprodução” (Barbosa, P., 2018, p. 45).

²⁹ O caso envolveu a Rural Telephone Service Company, empresa responsável pela publicação de listas telefônicas na região do Kansas, EUA, e a Feist Publications, Inc., editora especializada na distribuição de diretórios de telefones com uma cobertura maior do que a Rural. O caso se iniciou quando, diante de uma recusa da Rural em licenciar suas listas de páginas brancas para sua inclusão em um diretório que abrangeria 11 áreas diferentes de serviço telefônico, a Feist extraiu as listagens necessárias do diretório da Rural sem o consentimento do titular, e as publicou de forma idêntica à utilizada nas páginas brancas da Rural. O caso chegou à Suprema Corte, que

1.3.1. A originalidade nos dados e repositórios

Feitas estas considerações, passamos à seguinte dúvida: até que ponto os bancos de dados e, especialmente, o seu conteúdo, cumprem este requisito na prática? A resposta para esta pergunta necessita ser dividida em duas partes: primeiro, em relação aos repositórios e, em seguida, aos dados em si.

No tocante aos bancos de dados, embora seja possível conceber estruturas que reflitam escolhas do autor ou que fujam do lugar-comum, Derclaye (2008, p. 45) e Gervais (2019, p. 27-28) apontam que este não é o caso de boa parte dos acervos digitais. Gervais (2019, p. 27-28) destaca que se, à época do Acordo TRIPS, os formatos dos bancos de dados seguiam um formato relacional com índices ou tabelas que poderiam manifestar algum diferencial, isto mudou na última década, onde modelos não-relacionais, que sequer organizam arquivos, têm ganho popularidade. Mais especificamente, Banterle (2018, p. 5) aponta que, na maioria dos casos, os dados são coletados de forma automática, sem muita preocupação com a sua organização dentro da base de dados – porque tecnologias mais recentes permitem o processamento desses dados independentemente de como estiverem organizados.

No campo específico da mineração de textos, Martens (2018, p. 9) afirma que, para além da estruturação dos repositórios, tampouco parece haver preocupação com um processo de coleta e seleção “diferenciadas”: a extração de material segue uma lógica maximalista de obter o maior *corpus* de dados possível para análise, especialmente se a análise será feita com o auxílio de um algoritmo ou uma inteligência artificial. Afinal, tais ferramentas necessitariam de muito mais observações para realizar uma inferência que uma pessoa comum seria capaz de fazer com muito menos material à sua disposição. Por outro lado, segundo Truyens e van Eecke (2014, p. 163-164), sequer costuma ser tecnicamente vantajoso modificar deliberadamente o processo de coleta e organização, uma vez que questões de acessibilidade e compatibilidade terminam por obrigar os técnicos a adotar padrões já estabelecidos de montagem de bases de dados. Em outras palavras, o valor da maioria das bases de dados também depende de diversas restrições técnicas que geralmente inviabilizam o nível de escolha autoral ou não-banalidade necessários para que se alcance a originalidade.

Mas e no que se refere ao *conteúdo* da base de dados propriamente dito? Em outras palavras, é possível conceber que dados em si possam ser originais perante o Direito Autoral?

considerou que o conteúdo utilizado não satisfazia os requisitos de originalidade para a proteção autoral (ESTADOS UNIDOS DA AMÉRICA, 1991).

No Brasil, a legislação não é totalmente clara em relação a este tópico: a LDA não protege dados explicitamente no art. 7, embora tenha incluído as bases de dados como material protegido. Ao contrário, o § 2º do mesmo artigo, conforme já visto, não estende a proteção aos “dados ou materiais em si mesmos”. Por outro lado, a lista de obras protegidas pela lei brasileira não é exaustiva, aceitando criações “expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro”, incluindo “tais como” como evidência do caráter exemplificativo do dispositivo. Isto tudo é dificultado pela própria redação da lei, cujo art. 7, § 2º, destaca que a proteção da base de dados se dará “sem prejuízo de quaisquer *direitos autorais que subsistam a respeito dos dados ou materiais contidos nas obras*” (destaque nosso). Tal afirmação poderia ser interpretada de forma a autorizar que dados contidos em obras artísticas podem ser protegidos.

Uma abordagem mais certa, contudo, parece ser a mais simples e, ao mesmo tempo, a que comporta maior nuance interpretativa: observar o grau de originalidade dos dados em análise – e se o contexto no qual são mobilizados é capaz de evocar um elemento expressivo que seja original, ou não-banal. Afinal, assim como a legislação sugere que nem toda compilação de dados é considerada uma obra intelectual, é razoável presumir que o mesmo raciocínio deve se aplicar ao seu conteúdo.

No âmbito internacional, isto coincide, em alguma medida, com o entendimento alcançado na Organização Mundial de Propriedade Intelectual, na Segunda Sessão sobre Propriedade Intelectual e Inteligência Artificial (OMPI, 2020), de que (i) dados seriam uma forma de expressão sujeita ao teste de originalidade; (ii) se um dado representa uma obra original, ele será protegido, e (iii) “dados subsistentes em obras” são dados que representam obras protegidas.³⁰ Trata-se de uma observação coerente com a interpretação mais ampla de Borgman (2011): se entendermos livros, imagens, áudio, obras audiovisuais e artefatos do gênero como dados em si, parece natural presumir que estes dados se confundem com obras protegidas na sua totalidade.

Por outro lado, embora a submissão da expressão dos dados ao teste de originalidade pareça razoável, este trabalho adota uma posição divergente do entendimento de “dados em obras” acima em dois pontos. O primeiro deles diz respeito à própria noção de dados como construtos abstratos: se estes não existem de forma autônoma, necessitando ser materializadas

³⁰ ““Data subsisting in copyright works” is a term intended to distinguish between mere ideas which are not copyright protected and expressions of ideas which are protected. Data represent a mode of expression and data that represent original literary and artistic works are protected by copyright, also referred to “data subsisting in copyright works”.” (OMPI, 2020).

por meio de uma representação expressa (ou gráfica) (Garvey, 2013, p. 6-12), então seria razoável supor que este caráter abstrato tornaria os dados enquanto material que não constitui expressão, mas sim algo mais próximo de ideias – e, portanto, incapazes de receber proteção (Banterle, 2018, p. 4; Hugenholtz, 2017, p. 69), por força do princípio da dicotomia ideia/expressão: o Direito Autoral protege apenas a expressão, enquanto o conteúdo ou ideia manifesto nessa expressão deve permanecer livre. No direito brasileiro, isto é corroborado pelo art. 8º, I da Lei de Direito Autoral.³¹

Nosso segundo ponto de discordância está na possível interpretação de que “dados subsistentes em obras” implicam em dizer que dados contidos em obras, por definição, seriam originais, quando, na verdade, muitas expressões de dados existentes em uma obra (como palavras isoladas, ou um pixel de determinada cor em determinada posição) podem configurar dados “brutos”, de base factual. Em outras palavras, ainda que o entendimento de dados como abstrações sugira a presença de um certo grau de pessoalidade na sua construção (Garvey, 2013, p. 6-12), esta se encontra, de forma geral, geralmente conscrita a um objetivo prático: representar características ou atributos de entidades reais. Nesse caso, ainda que existam em um plano distinto dos fatos em si, é de se questionar se os dados – especialmente os dados “brutos”, como os metadados – não poderiam ser equiparados a fatos – o que ressoa com a noção de veracidade dos megadados, onde seu valor está intrinsecamente ligado à sua correspondência aos fatos.

A razão para este questionamento se dá na presença constante, no Direito Autoral, de um rechaço institucional à ideia de que fatos propriamente ditos sejam protegidos para além da expressão artístico-literária que os retrata. Um exemplo disso pode ser visto no Artigo 2 (8) da Convenção de Berna, segundo o qual não são obras protegidas as notícias do dia ou similares que possuam o caráter de mera informação para fins de imprensa.³² A LDA, embora não exclua fatos explicitamente da proteção do Direito Autoral, também provê exemplos que sugerem essa proibição: o art. 7º, que protege a expressão artística ou literária do texto acadêmico, mas não cobre seu conteúdo – ou, mais especificamente, a informação ou os dados ali existentes.³³ Outro

³¹ “Art. 8º Não são objeto de proteção como direitos autorais de que trata esta Lei:

I - as idéias, procedimentos normativos, sistemas, métodos, projetos ou conceitos matemáticos como tais;” (Brasil, 1998b).

³² “Article 2 (...) (8) (8) La protection de la présente Convention ne s’applique pas aux nouvelles du jour ou aux faits divers qui ont le caractère de simples informations de presse.” (OMPI, 1979). Aqui, optou-se pela citação original em francês, uma vez que a versão em inglês utiliza a expressão “miscellaneous facts” (OMPI, 1979) (fatos avulsos) como tradução literal do termo “faits divers” (“notícias”, em francês), causando a impressão de que o artigo menciona fatos de forma explícita. Ainda assim, argumentamos que a ideia central de que a Convenção afasta a proteção a fatos ainda pode ser deduzida do texto original.

³³ Art. 7º

(...)

exemplo é, novamente, o art. 8º – que exclui explicitamente da proteção do instituto as informações de uso comum³⁴ – se entendermos que o uso comum de um dado isolado é representar ou descrever uma entidade, ação ou fenômeno (Rowley, 2007; Kelleher e Tierney, 2018).

Indo mais além, este trabalho argumenta que tal tendência encontra paralelos em outros campos da propriedade intelectual: na Lei nº 9.279/96 (a Lei de Propriedade Industrial brasileira), por exemplo, não se consideras patenteáveis as “descobertas, teorias científicas e métodos matemáticos”, as “concepções puramente abstratas” ou elementos encontrados na natureza, incluindo o seu genoma (Brasil, 1996);³⁵ tampouco podem ser considerados como marca quaisquer símbolos ou termos que sejam meramente descritivos ou designadores de um bem ou serviço, tampouco formatos que sejam indissociáveis de uma funcionalidade técnica.³⁶

A jurisprudência brasileira nos fornece um entendimento similar, especialmente no que concerne aos dados científicos. Nesse sentido, duas decisões merecem destaque: a primeira, que antecede a própria LDA, data de 25/05/1979, quando o Supremo Tribunal Federal julgou o Recurso Extraordinário no. 88705/RJ, que tratava da reprodução de conclusões e ensinamentos científicos em uma bula de remédio. O Tribunal, ali, entendeu que tal conteúdo pertence ao bem comum, não integrando, portanto, a esfera de proteção do Direito Autoral – que tão somente protege a forma pela qual tais ensinamentos foram expressos.³⁷ O segundo caso de destaque é o REsp 1528627/SC, julgado pelo Superior Tribunal de Justiça em 07/03/2017, o qual

§ 3º No domínio das ciências, a proteção recairá sobre a forma literária ou artística, não abrangendo o seu conteúdo científico ou técnico, sem prejuízo dos direitos que protegem os demais campos da propriedade imaterial.” (Brasil, 1998b).

³⁴ “Art. 8º Não são objeto de proteção como direitos autorais de que trata esta Lei: (...)

V - as informações de uso comum tais como calendários, agendas, cadastros ou legendas; (...).” (Brasil, 1998b).

³⁵ Ver o art. 10, I, II e IX da Lei nº 9.279/96 (Brasil, 1996).

³⁶ “Art. 124. Não são registráveis como marca:

(...)

VI - sinal de caráter genérico, necessário, comum, vulgar ou simplesmente descritivo, quando tiver relação com o produto ou serviço a distinguir, ou aquele empregado comumente para designar uma característica do produto ou serviço, quanto à natureza, nacionalidade, peso, valor, qualidade e época de produção ou de prestação do serviço, salvo quando revestidos de suficiente forma distintiva;

(...)

XXI - a forma necessária, comum ou vulgar do produto ou de acondicionamento, ou, ainda, aquela que não possa ser dissociada de efeito técnico;” (Brasil, 1996).

³⁷ “EMENTA: -Bula de remédios. Reconhecida a sua natureza científica, pois destinada à classe médica e farmacêutica e fiscalizada pelas autoridades competentes, legitima-se a simples referência ou citação de uma pesquisa científica sem ofensa ao direito do autor, face ao Código Civil e à lei especial.

Aplicação dos arts. 666, I, do C. Civil e art. 49, I e III da Lei 5.988, de 14.02.73, e não incidência, na espécie, do art. 23, IV, da mesma lei.

Aos trabalhos científicos o direito autoral protege a forma de expressão, e não as conclusões científicas ou seus ensinamentos, que pertencem a todos, no interesse do bem comum.

Ação proibitória cujos resultados já tenham sido alcançados pela notificação prévia.

RE conhecido e provido para julgar improcedente a ação proibitória. (RE n. 88.705-9/RJ, Ministro Cordeiro Guerra, Segunda Turma, julgado em 25/5/1979).” (Brasil, 1979. Destaque nosso). Disponível em <https://re-dir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=182506>. Acesso em 22 fev. 2024.

determinou que a proteção conferida ao texto de uma dissertação de mestrado não se estendia às ideias ou à bibliografia ali contidas.³⁸ Fora do Brasil, também há exemplos similares na *common law*, a exemplo dos Estados Unidos: o caso *Feist* classifica dados “brutos” como equivalentes a fatos e, portanto, como extrínsecos ao escopo do Direito Autoral, uma vez que existem de forma independente do autor.³⁹

Portanto, parece ser mais acertado entender que a maioria dos dados dificilmente configuraria material protegido: não apenas porque a proteção recairia não sobre eles, e sim sobre sua expressão em um meio específico (ver o art. 7º, *caput*, da LDA), mas, principalmente, porque um dado isolado não é inteligível o suficiente para fazer sentido, (Rowley, 2007), muito menos para constituir material original: com isto em mente, e tomando-se como base a ideia de que originalidade seria um limite mínimo a partir do qual uma obra sai do banal, do lugar-comum (Barbosa, D., 2005; Barbosa, P., 2018), então os dados existentes em uma obra só podem gerar expressão original se existirem em um número suficiente e se forem relacionados em uma forma suficientemente coesa e compreensível enquanto obra.

1.3.2. Mineração de textos e dados e sua licitude no Direito Autoral

Constatações tais como as vistas acima se tornam importantes quando analisamos o status de atividades de coleta e análise de dados quando estas fazem uso de obras protegidas:

³⁸ “RECURSO ESPECIAL. DIREITO CIVIL. PROPRIEDADE INTELECTUAL. DIREITOS AUTORAIS. ALEGAÇÃO DE OFENSA A DIREITOS AUTORAIS NA CRIAÇÃO DE DISCIPLINA DE PÓS-GRADUAÇÃO DA FACULDADE DE LETRAS DA UFSC. OBJETIVOS DA DISCIPLINA E BIBLIOGRAFIA BÁSICA QUE SERIAM IDÊNTICOS AO DE DISSERTAÇÃO DE MESTRADO. NÃO CONFIGURAÇÃO DA ALEGADA OFENSA. PROTEÇÃO DOS DIREITOS AUTORAIS QUE NÃO TÊM A EXTENSÃO PRETENDIDA PELA AUTORA.

1. Pretensão da autora de condenar a Universidade Federal de Santa Catarina a retirar do programa de pós-graduação da Faculdade de Letras disciplina relativa à história literária dos vampiros, que teria sido criada com base em sua dissertação de mestrado.

Pretensão, também, de recebimento de indenização por danos extrapatrimoniais.

2. Embora o texto de sua dissertação e das suas conferências realizadas em eventos acadêmicos esteja protegido pela Lei de Direitos Autorais, as ideias que lhe serviram de base, bem como a bibliografia de que se valeu para a realização da pesquisa, não estão abarcadas pela proteção aos direitos de autor.

3. Inteligência dos arts. 7º, parágrafos 2º e 3º, e 8º da Lei n. 9.610/98.

4. Pretensão recursal que esbarra, ainda, na Súmula 7/STJ, pois exigiria o reexame de matéria fático-probatória.

5. RECURSO ESPECIAL DESPROVIDO.

(Brasil, 2017).

³⁹ Ver *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 111 S. Ct. 1282, 1296, 113 L. Ed. 2d 358 (1991) (“In other words, did Feist, by taking 1,309 names, towns, and telephone numbers from Rural's white pages, copy anything that was “original” to Rural? Certainly, the raw data does not satisfy the originality requirement. Rural may have been the first to discover and report the names, towns, and telephone numbers of its subscribers, but this data does not “‘ow[e] its origin’” to Rural. *Burrow–Giles*, 111 U.S., at 58, 4 S.Ct., at 281. Rather, these bits of information are uncopyrightable facts; they existed before Rural reported them and would have continued to exist if Rural had never published a telephone directory. The originality requirement “rule[s] out protecting ... names, addresses, and telephone numbers of which the plaintiff by no stretch of the imagination could be called the author.” *Patterson & Joyce 776*”).

afinal, se o material efetivamente coletado e analisado é equiparável a ideias ou fatos e não atinge esta massa crítica de quantidade ou inteligibilidade enquanto expressão original, isto significaria que a mineração de textos e dados, a princípio, poderia ser considerada lícita perante a legislação autoralista, na medida em que o seu propósito seja, mais especificamente, fazer sentido dos dados enquanto informação.

Nesse sentido, alguns autores, a exemplo de Sag (2012), Grimmelman (2015) e Carroll (2020), identificaram uma doutrina jurídica que está intrinsecamente ligada a esta delimitação de materiais puramente factuais como externos ao Direito Autoral. Trata-se da noção de “obras não-expressivas” – aquelas baseadas puramente em material factual, como listas telefônicas – e de “uso não-expressivo”, aqui entendido como qualquer uso que não tenha como propósito explorar os aspectos expressivos de uma dada obra. Por exemplo, um uso expressivo envolveria “baixar” um filme na internet ou fotocopiar um livro para consumo posterior, enquanto outras atividades, como contagem de palavras ou extração de metadados, constituiriam usos não-expressivos, sobre os quais os direitos exclusivos dos titulares não incidiriam. No âmbito regulatório da União Europeia, a European Internet Society (2017) também manifestou a opinião de que os processos tecnológicos envolvidos na mineração de textos e dados meramente apreenderiam dados e outros elementos não-protegidos das obras porventura utilizadas – sobre as quais não incidem direitos autorais.

Por outro lado, Sobel (2017) apresenta um exemplo contrário a este movimento, apontando para um progressivo crescimento da quantidade de sistemas de inteligência artificial capazes de absorver mais do que simples conteúdo factual. Desta maneira, o autor levanta a possibilidade de que o nível de granularidade e detecção de padrões em sistemas mais recentes confira à diversos modelos a capacidade de identificar conjuntos de atributos que vão além de ideias ou fatos, identificando como autores manifestam essas ideias – algo muito mais próximo do que se consideraria expressão protegida.⁴⁰

⁴⁰ “By uncovering “interesting structure” in data rather than emphasizing features that humans have predetermined to be salient, machine learning techniques can extract and mimic features that people might find ineffable or difficult to discern, but that nevertheless encapsulate some unifying quality of a particular group of data. When trained on samples of an individual’s handwriting or recordings of a human voice, for example, machine learning models can mimic scrawls and drawls with uncanny accuracy.⁷⁹ The constellations of features that machine learning can appropriate might well be called “personality.” This capability complicates a prevailing assumption in copyright law: that traces of an author’s “personality” uniquely individuate works of authorship, and that those traces ought to be copyrightable.

(...)

(R)ecent advances in machine learning may refute the idea that unique personality subsists only in the individual human gesture, or at least undermine it as a justification of intellectual property rights. Machine learning techniques enable machines to identify and mimic the features that distinguish sensory data, even when those features are not qualities that humans can easily express or represent. Today’s technology can isolate the characteristics

Já Craig apresenta um posicionamento com outra nuance importante: por um lado, a autora (Craig, 2022) rejeita a noção que os sistemas do gênero vistos até então seriam dotados de capacidade cognitiva e intencionalidade suficiente para conceber e processar obras mais do que como amontoados de dados e processos de adivinhação – o que lhes negaria a condição de autor, mas também lhes colocaria fora da posição de infringentes.⁴¹ Por outro lado, em outro texto (Craig, 2021), afirma que isto não muda o fato de que, se um processo de mineração para fins de treinamento de uma IA utilizou uma base de dados que continha uma obra protegida e produz um resultado similar àquela obra, é possível argumentar pela presença de elementos jurídicos suficientes para constatar uma violação de direitos autorais – o que, embora complicado pela falta de responsabilidade jurídica de uma máquina, não exclui necessariamente a responsabilidade dos humanos envolvidos na sua operação.⁴²

Esta ênfase não no processo, mas no resultado, ganha relevância quando se observa que, se as fontes de uma dada coleta e análise de dados envolverem mais do que dados “brutos” ou factuais e incluírem obras protegidas em seu acervo, então as operações técnicas necessárias à mineração de textos e dados implicam em usos que dependem de autorização dos titulares das obras. Mais especificamente, procedimentos dessa natureza que utilizam textos, imagens ou outro tipo de material protegido envolvem a reprodução, adaptação e, por vezes, comunicação ao público do conteúdo, ainda que em parte, das obras coletadas (Geiger, Frosio e Bulayenko, 2018; Instituto Brasileiro de Direitos Autorais, 2024). Tal posicionamento também é confirmado em Nota Técnica recentemente emitida pela Secretaria de Direitos Autorais do Ministério da Cultura (Brasil, 2024a, p. 9-11).

Há ainda um elemento adicional, peculiar ao Brasil: segundo o art. 29, IX da LDA, a mera inclusão de uma obra em uma base de dados já constitui ato que depende de prévia e

that individuate a human face and human handwriting, the very attributes that Hargrave and Justice Holmes used long ago as metonyms for protectable authorial personality.” (Sobel, 2017, p. 59-60).

⁴¹ “Authors deem themselves to be authoring, and in doing so, their intention is to “produce mental effects” in their audience—they intend to “generate thoughts, feelings, emotions, and other states of cognition.” If this is our measure, once again, AI is incapable of the requisite intentionality. (...)The AI is not an audience capable of such a response, of course, and so the use of the work to train the AI amounts only to use of it as a functional thing and as a source of information or data (which belongs in the public domain). It is not a use of the work *as a work of authorship*. It should not, therefore, implicate copyright at all.” (Craig, 2022, p. 147, 152-153).

⁴² “Where a substantial similarity between an AI output and a protected work is simply a matter of coincidence (as where Basanta’s scanners independently captured images without copying) the similarity is no basis for liability. However, if an AI is trained on a dataset that includes a particular protected work and subsequently produces a substantially similar output, then it is harder to chalk the similarity up to coincidence: there is access to the protected work and so the necessary causal connection. (...) But even so, who is to be sued? Applying a principal-agent analysis, the machine’s volitional act *qua* “agent” could potentially produce liability for whichever party notionally fits the role of “principal”—most likely the person that creates or deploys the AI.” (Craig, 2021, p. 14-15).

expressa autorização do titular.⁴³ Isto significa que, perante o Direito Autoral brasileiro, sequer é necessário fazer um exercício mental para entender a existência de reprodução ou adaptação: proíbe-se o próprio armazenamento do material sem autorização prévia e expressa do seu titular. Ora, se, como visto anteriormente, tal operação constitui uma etapa corriqueira e necessária do processo de análise de dados – afinal, não se pode analisar um material que não foi coletado –, então, o dispositivo cobriria mais do que afirma explicitamente, efetivamente impedindo que processos de mineração de textos e dados de uma forma geral possam ser conduzidos caso utilizem fontes que incluam no mínimo uma única obra artística ou literária em seu meio.

Isto também se aplicaria não só a obras convencionais em si, mas também a uma base de dados original, caso esta seja utilizada como fonte para a coleta e análise mesmo que o conteúdo deste repositório não seja protegido: há potencial violação de direitos de reprodução da estrutura da base (especialmente se for copiada uma parte substancial do acervo), bem como ofensa a direitos de adaptação caso seja feito o descarte de conteúdo irrelevante dessa base, de forma a alterar a estrutura da base original, ou a direitos de divulgação (Geiger, Frosio e Bulayenko, 2018, p. 6-7). No Brasil, isto encontra respaldo no art. 87 da LDA.⁴⁴

Isto suscita uma questão: ainda que a lei expressamente afirme que a proteção de uma base de dados original seja separada da estendida ao seu conteúdo, em que ponto a reprodução de um conteúdo termina e a reprodução da estrutura da base começa? O que separa a expressão da base com a dos elementos ali inseridos? A falta de uma definição concreta para estes requisitos, argumentamos, abre brecha para o uso de repositórios como mecanismo de impedimento para o uso de seu conteúdo, mesmo que este seja composto por obras em domínio público ou por materiais sobre os quais não recai direito de autor.

⁴³ “Art. 29. Depende de autorização prévia e expressa do autor a utilização da obra, por quaisquer modalidades, tais como:

I - a reprodução parcial ou integral;

II - a edição;

III - a adaptação, o arranjo musical e quaisquer outras transformações;

IV - a tradução para qualquer idioma;

(...)

IX - a inclusão em base de dados, o armazenamento em computador, a microfilmagem e as demais formas de arquivamento do gênero; (...).” (Brasil, 1998b).

⁴⁴ “Art. 87. O titular do direito patrimonial sobre uma base de dados terá o direito exclusivo, a respeito da forma de expressão da estrutura da referida base, de autorizar ou proibir:

I - sua reprodução total ou parcial, por qualquer meio ou processo;

II - sua tradução, adaptação, reordenação ou qualquer outra modificação;

III - a distribuição do original ou cópias da base de dados ou a sua comunicação ao público;

IV - a reprodução, distribuição ou comunicação ao público dos resultados das operações mencionadas no inciso II deste artigo.” (Brasil, 1998b).

A título de ilustração, consideremos a base de dados WIPOLex,⁴⁵ operada pela Organização Mundial da Propriedade Intelectual (OMPI). Trata-se de um repositório destinado ao fornecimento de informações públicas acerca da proteção à propriedade intelectual (direito de autor incluso) de todos os seus Países-Membros. Nota-se que os termos de uso da plataforma permitem a reprodução para fins de pesquisa e outros não-comerciais, mas proíbem diversos usos de interesse para a mineração de textos e dados: por exemplo, a possibilidade de se extrair, armazenar e utilizar grandes conjuntos de dados ali contidos, ou realizar *web scraping*.⁴⁶ Ainda que estes termos sejam de ordem contratual, consideremos a possibilidade de esta base ser considerada original: nesses termos, a proibição do uso de ferramentas de *scraping* não faz diferenciação quanto ao tipo de conteúdo: em tese, não se pode minerar os textos legais, de domínio público, que compõem a base, o que constitui uma clara expansão da proteção que seria possível pelo Direito Autoral.

Ao mesmo tempo, é importante observar que diversos sistemas mais avançados de bases de dados e de inteligência artificial parecem atuar de uma maneira que desafia a lógica exposta acima: embora alguns dos modelos hoje operantes utilizem repositórios próprios, diversos deles explicitamente utilizam, majoritariamente, fontes de terceiros. Ademais, diversos modelos já não necessitam manter cópias do material obtido; por questões de economia de espaço, apenas mantêm representações compactadas de elementos estatísticos presentes nos dados, que existem em um formato aglomerado que não pode ser diretamente ligado a nenhuma obra em particular.⁴⁷ É o caso do LAION-5B, que utiliza bilhões de imagens para a construção

⁴⁵ Disponível em: <https://www.wipo.int/en/web/wipolex/index>. Acesso em 10 set. 2024.

⁴⁶ “2. Use of data

2.1 The WIPO Lex Database is a free-of-charge public service aimed at implementing WIPO's objective to assemble and disseminate information concerning the protection of intellectual property, as set forth in the Article 4(vi) of the Convention Establishing WIPO. In view of this objective, the WIPO Lex Database is provided for public information purposes and WIPO allows the User to access and make use of the information it contains under the conditions described herein.

2.1.1 IP Legal Texts and Translated Texts

The User is allowed to reproduce the IP Legal Texts and Translated Texts for academic, research and non-commercial purposes, provided that the original source and "WIPO Lex" are acknowledged.

(...)

2.2 Any abusive use of the WIPO Lex Database is prohibited and the User is therefore forbidden to:

(...)

c. perform bulk acquisition, bulk downloading and bulk storing of data;

d. perform bulk copying, bulk reformatting, bulk sharing and bulk redistributing of data;

e. perform web scraping;” (OMPI, 2024).

⁴⁷ “There is a common misconception that a model is just a dataset that contains a copy of every single entry in the dataset. (...) The trained model doesn't contain a copy of the books, in danger of over-simplifying things, such a model contains statistics, the likelihood of a token following another.(...) (T)he takeaway is that models do not contain copies of works. For legal purposes, they are not even derivatives of one specific work in the dataset, the idea of large datasets is precisely that a model is not based on one individual work. This is evidenced by the size of trained models, while some are relatively large, they run in gigabytes, not tera or petabytes, and cannot possibly contain all the works in the training data, they are mostly statistical data.” (Guadamuz, 2024, p. 8)

de seu sistema, mas mantém apenas links e metadados sobre as mesmas em seu repositório (Beaumont, 2022).

Também é importante destacar que a proteção explícita sobre os usos dessas obras vista acima não é o único componente à disposição para impedir o uso não-autorizado de obras: como veremos a seguir, para além do caráter inerentemente reativo da lei – que não impede infrações, mas as pune retroativamente (Derclaye, 2008) – o Direito Autoral é apenas um componente dentro de um sistema mais amplo de proteção formado para “cobrir todas as bases” dos titulares de direitos – inclusive apostando na tecnologia como resolução.

1.3.3. O sistema de camada tripla no Direito Autoral

Dusollier (2005) aponta que este sistema pode ser dividido em três camadas, cada uma sendo reforçada pela próxima: no primeiro nível, tem-se a proteção do material por Direito Autoral, já vista acima; no segundo, temos as barreiras tecnológicas e, no terceiro, há as disposições anti-burla.

No que tange à barreira tecnológica, esta é composta pelos mecanismos de Digital Rights Management (DRM) e pelas medidas tecnológicas de proteção (*technological protection measures*, ou TPM, que aqui chamaremos de MTP). Mais especificamente, os mecanismos de DRM são comumente usados para introduzir metadados às obras que facilitem a identificação dos titulares, enquanto as MTP servem para controlar e restringir o acesso e uso não-autorizado de obras por meios técnicos. No entanto, seu uso abusivo já foi denunciado por diversos autores: da imposição de cláusulas abusivas e restrição ao acesso e uso de dados não-protegidos ou obras em domínio público à instalação de programas espíões, aponta-se que esses mecanismos muitas vezes atuam *extra legem*, utilizando a arquitetura digital como efetivo mecanismo de regulação extralegal que permite o controle sobre acesso e uso de conteúdo protegido para além do escopo garantido pelos direitos autorais, restringindo usos legítimos de material licitamente adquirido (Lessig, 1999, p. 135-138, 2002; Lewicki, 2007, p. 228-229; Branco, 2011, p. 269; Souza, A. e Schirru, 2016, p. 41).⁴⁸

Já as medidas anti-burla (do inglês *anti-circumvention measures*) são dispositivos legais que vedam a alteração, supressão, modificação ou inutilização desses mecanismos tecnológicos (por exemplo, quando um processo de raspagem remove a marca d'água de uma imagem ou ultrapassa o *paywall* de um texto), bem como a distribuição ou comercialização de

⁴⁸ Para mais informações sobre a forma como a mecanismos digitais são aplicados como forma extralegal de regulação, ver Lessig (1999).

aparatos ou programas destinados a essas atividades. Na esfera internacional, tais mecanismos legais foram tornados oficiais com o Tratado da OMPI sobre Direito de Autor (OMPI, 1996).⁴⁹ No momento da conclusão desta tese, o Brasil não consta como signatário deste documento;⁵⁰ no entanto, o país incorporou uma medida do gênero no art. 107 da Lei nº 9.610/98.⁵¹ Tais provisões são postas em um terceiro nível por não visarem a proteção de obras de forma material, mas sim a proteção dos mecanismos tecnológicos de DRM e MTP. Configuram, portanto, uma espécie de “paracopyright” ou “super-copyright” que opera à margem do objeto central do Direito Autoral, mas ainda dentro do sistema (Brown, 2003; Derclaye, 2008, p. 196).

Na análise de Derclaye (2008, p. 196-197; 201-205), isto representa um problema quando se observa que, muitas das vezes, essas camadas adicionais não se submetem aos limites tradicionais do Direito Autoral – como, por exemplo, provisões legais que permitam o uso de obras sem autorização – ou, quando o fazem, é de forma limitada. Ademais, a autora lembra que contornar uma MTP muitas vezes constitui um ato além das capacidades técnicas do usuário

⁴⁹ “Article 11

Obligations concerning Technological Measures

Contracting Parties shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection with the exercise of their rights under this Treaty or the Berne Convention and that restrict acts, in respect of their works, which are not authorized by the authors concerned or permitted by law.

Article 12

Obligations concerning Rights Management Information

(1) Contracting Parties shall provide adequate and effective legal remedies against any person knowingly performing any of the following acts knowing, or with respect to civil remedies having reasonable grounds to know, that it will induce, enable, facilitate or conceal an infringement of any right covered by this Treaty or the Berne Convention:

- (i) to remove or alter any electronic rights management information without authority;
- (ii) to distribute, import for distribution, broadcast or communicate to the public, without authority, works or copies of works knowing that electronic rights management information has been removed or altered without authority.

(2) As used in this Article, “rights management information” means information which identifies the work, the author of the work, the owner of any right in the work, or information about the terms and conditions of use of the work, and any numbers or codes that represent such information, when any of these items of information is attached to a copy of a work or appears in connection with the communication of a work to the public.⁹

⁵⁰ A menção do texto do Tratado é mantida, contudo, não apenas em razão de o Brasil possuir um dispositivo que a “cumpra” de forma independente, mas também em função de importar pra a análise de modelos de regulação de Direito Autoral pertencentes a outros países – esses, sim, signatários do Tratado –, a exemplo dos questionamentos feitos por Gervais (2019), que servirão de base teórica em um momento posterior deste trabalho.

⁵¹ “Art. 107. Independentemente da perda dos equipamentos utilizados, responderá por perdas e danos, nunca inferiores ao valor que resultaria da aplicação do disposto no art. 103 e seu parágrafo único, quem:

- I - alterar, suprimir, modificar ou inutilizar, de qualquer maneira, dispositivos técnicos introduzidos nos exemplares das obras e produções protegidas para evitar ou restringir sua cópia;
- II - alterar, suprimir ou inutilizar, de qualquer maneira, os sinais codificados destinados a restringir a comunicação ao público de obras, produções ou emissões protegidas ou a evitar a sua cópia;
- III - suprimir ou alterar, sem autorização, qualquer informação sobre a gestão de direitos;
- IV - distribuir, importar para distribuição, emitir, comunicar ou puser à disposição do público, sem autorização, obras, interpretações ou execuções, exemplares de interpretações fixadas em fonogramas e emissões, sabendo que a informação sobre a gestão de direitos, sinais codificados e dispositivos técnicos foram suprimidos ou alterados sem autorização.” (Brasil, 1998b).

comum. Isto implica, efetivamente, que, se o titular de direitos impedir o acesso a um material sob a justificativa de prevenir sua cópia ou utilização sem autorização, então tem-se, na prática, um direito exclusivo sobre o acesso – estendendo o escopo de direitos exclusivos do titular para além dos limites legais, que deveriam ser relativos única e exclusivamente à exploração das obras.⁵²

No caso mais específico da mineração de textos e dados, isto tudo se torna mais agravante pelo fato de que esse tipo de processo (ou, no mínimo, a coleta e a transformação de conteúdo) geralmente exigem a cópia e a análise de uma grande quantidade de material, em sua totalidade - seja para garantir que não haja lacunas nos dados que possam levar a resultados incompletos ou incorretos (Levendowski, 2018), seja para treinar adequadamente os algoritmos para que eles sejam capazes de executar suas funções a contento.⁵³

Nesse cenário, qualquer projeto que se baseie em mineração de textos e dados e que não queira se sujeitar a possíveis desgostos judiciais se depara com duas opções: a primeira é evitar fontes proprietárias, o que pode comprometer os resultados da pesquisa devido a deficiências no material separado para análise. A segunda é entrar em contato com os detentores dos direitos sobre as obras e os bancos de dados envolvidos e obter autorização de uso, o que pode ser uma tarefa extremamente demorada e cara. Afinal, embora a busca por todos os detentores de direitos autorais relevantes esteja longe de ser um problema novo no licenciamento de direitos autorais,⁵⁴ isso se torna ainda maior com o grande número de dados necessários em seu TDM típico, o que pode implicar em um grande número de detentores diferentes (Martens, 2018, p. 9).

Reichman e Okediji (2012) mencionam a proteção dos bancos de dados como parte das novas barreiras para os pesquisadores que desejam lidar com técnicas automatizadas de

⁵² “(...) (P)rotection against circumvention is very broad as it encompasses any act not authorized by the right holder. This includes the act of accessing works (...). Thus if the right Holder does not authorize access, consultation or use of their database, or acts of reutilization after the expiration of the term of protection for the purposes criticism or review or reporting current events, or does not authorize transfer, then it is illegal to circumvent the TPM to perform such acts (...). TPMs and anti-circumvention provisions should not in effect grant rights of access and use to database producers. Rights to prevent the access and use of a database are rights additional to the rights of extraction and reutilization (...). Right holders cannot privately create and enforce those rights. This seems to have been acknowledged implicitly by the ECJ (...) as it ruled that rights of extraction and reutilization do not cover consultation of the database” (Derclaye, 2008, p. 196-197, 205).

⁵³ Ver Martens (2018, p. 9) (“[machine learning] não só é capaz de lidar com conjuntos de dados muito maiores do que os humanos podem lidar; ele também requer conjuntos de dados muito maiores do que os humanos precisam para aprender. Enquanto os seres humanos podem exigir apenas algumas observações para aprender uma resposta comportamental, o AM pode exigir milhares ou milhões de observações para aprender algumas respostas básicas”) (tradução nossa).

⁵⁴ Para um exemplo mais detalhado dos obstáculos envolvidos no licenciamento de direitos autorais quando várias obras estão envolvidas, ver Lessig (2004).

"descoberta de conhecimento" e o papel dos periódicos e dos financiadores nesse processo. Ducato e Stowel apontam como as restrições de propriedade intelectual em relação ao uso de dados têm operado em relações contratuais em plataformas digitais e como elas interferem em certas tecnologias benéficas para o usuário (por exemplo, programas que simplificam os termos de uso para leigos). Levendowski, por outro lado, reforça que a preferência pelo uso de repositórios de "baixo risco" para treinamento de algoritmos tende a potencializar a exacerbação de resultados tendenciosos em sistemas de IA, em razão da baixa quantidade ou qualidade dos dados utilizados. Outros, como Hugenholtz (2017) e Geiger, Frosio e Bulayenko (2018), afirmam que uma proteção excessivamente ampla contribuiria para reduzir a acessibilidade dos dados públicos, gerando um clima de incerteza regulatória para aqueles que trabalham com dados.

1.4. O direito *sui generis*

Neste ponto, uma breve pausa se faz necessária para abordar outro domínio da propriedade intelectual que, embora não diga respeito diretamente ao Direito Autoral, tem imediata relação com o tema. Trata-se da proteção *sui generis* ao conteúdo de bases de dados não-originais.

Conforme visto acima, bancos de dados só recebem proteção pelo Direito Autoral na medida em que a seleção, organização ou arranjo do seu conteúdo constituam expressão original – e não o conteúdo em si. Também se observou que muitas bases não satisfazem esse critério por razões de ordem técnica ou prática, o que exclui um grande número de materiais.

Contudo, a fim de proteger investimentos consideráveis em recursos humanos, técnicos e financeiros empreendidos na construção de bases de dados e impedir que a extração ou uso não autorizados do total ou de parte substancial do conteúdo dessas bases desincentivem a criação de repositórios e obstaculizem o desenvolvimento de um mercado de informação dentro da União Europeia, o bloco implementou a Diretiva 96/9/CE (União Europeia, 1996, p. 1)⁵⁵,

Tal Diretiva visa não apenas reiterar a proteção a bases de dados originais pelo Direito Autoral, como criar um direito *sui generis* que garante, aos produtores de bancos de dados digitais não-originais (ou seja, que não são protegidos pelo Direito Autoral) o direito de exclusividade sobre a extração e reutilização total ou substancial do conteúdo dos seus repositórios. Aqui, o critério essencial para proteção não é nenhuma consideração criativa, sendo a

⁵⁵ “Considerando que o fabrico de uma base de dados exige o investimento de recursos humanos, técnicos e financeiros consideráveis, podendo-se copiar ou aceder a essas bases a um custo muito inferior ao de uma concepção autónoma de uma base de dados (...)” (União Europeia, 1996, p. 1).

justificativa uma questão de investimento: protege-se o conteúdo (e não a estrutura) de uma base não-original caso “a obtenção, verificação ou apresentação desse conteúdo representem um investimento substancial do ponto de vista qualitativo ou quantitativo” (União Europeia, 1996, p. 6).⁵⁶

Tal direito tem duração distinta daquela concedida à estrutura de bases de dados originais: o Artigo 10 da Diretiva confere um prazo de 15 anos, contados a partir de 1º de janeiro do ano subsequente ao de sua conclusão, podendo ser renovado por mais 15 anos se, antes do seu término, a base for posta à disposição do público. Ademais, adições, supressões ou alterações sucessivos que levem à configuração de um novo investimento substancial podem ensejar um novo prazo de proteção (União Europeia, 1996, p. 7).⁵⁷ Na prática, isto potencialmente levaria a um direito perpétuo, uma vez que não fica claro como medir até que ponto houve novo investimento substancial, de forma que há abertura para que se renove a exclusividade de uso de um dado repositório em um número indefinido de vezes (Derclaye, 2008).

Isto se agrava ainda mais com a realidade dos megadados: como visto anteriormente, enormes quantidades de dados são produzidas a todo momento – o que leva a uma lógica de que o conteúdo de qualquer repositório que seja mantido com regularidade terá consideráveis transformações em pouco tempo. Sem um critério definido para a relevância de um investimento, dá-se abertura para uma fácil constatação de que uma renovação de proteção seria cabível. Contudo, cabe destacar que há algumas restrições ao escopo do tipo de material coberto pelo direito *sui generis*, como será visto no Capítulo 3.

Este tipo de regulação já encontrou iniciativas de replicação para fora da União Europeia. Um exemplo disso está na Coreia do Sul, que também possui uma regulação de

⁵⁶ “Artigo 7º

Objecto da protecção

1. Os Estados-membros instituirão o direito de o fabricante de uma base de dados proibir a extracção e/ou a reutilização da totalidade ou de uma parte substancial, avaliada qualitativa ou quantitativamente, do conteúdo desta, quando a obtenção, verificação ou apresentação desse conteúdo representem um investimento substancial do ponto de vista qualitativo ou quantitativo” (União Europeia, 1996, p. 6).

⁵⁷ “Artigo 10º

Prazo de protecção

1. O direito previsto no artigo 7º produz efeitos a partir da data de conclusão do fabrico da base de dados, e expira ao fim de 15 anos a contar de 1 de Janeiro do ano seguinte ao da data de conclusão.

2. No caso de uma base de dados que tenha sido posta à disposição do público antes do decurso do prazo previsto no nº 1, o prazo de protecção por este direito terminará ao fim de quinze anos a contar de 1 de Janeiro do ano seguinte àquele em que a base de dados tiver sido posta pela primeira vez à disposição do público.

3. Qualquer modificação substancial, avaliada quantitativa ou qualitativamente, do conteúdo de uma base de dados, incluindo quaisquer modificações substanciais resultantes da acumulação de aditamentos, supressões ou alterações sucessivos que levem a considerar que se trata de um (sic) novo investimento substancial, avaliado qualitativa ou quantitativamente, permitirá atribuir à base resultante desse investimento um período de protecção próprio” (União Europeia, 1996, p. 7).

propriedade para bancos de dados não-originais – inclusive com uma limitação similar para fins de pesquisa, mas com um prazo de proteção menor (cinco anos).⁵⁸

A possibilidade da incorporação deste tipo de regulação a nível de tratados internacionais chegou a ser discutida na Organização Mundial da Propriedade Intelectual, que levantou estudos a respeito. Os resultados foram mistos: por exemplo, Braunstein (2002, p. 27) aponta para a adoção do direito *sui generis* por países em desenvolvimento, com poucas limitações ou exceções, enquanto medida para estimular e proteger o investimento em bases de dados nesses locais. Por outro lado, o estudo de López (2002, p. 30-31) expressa o receio de que a adoção de uma regulação do tipo em âmbito internacional não se adequaria ao arcabouço jurídico da maioria dos países, além de potencialmente trazer prejuízos para a circulação do conhecimento e para o desenvolvimento econômico no Sul Global. No final das contas, a proposta nunca chegou a seguir adiante, com forte oposição de diversos países (Davison 2003, p. 230-234).

⁵⁸ “Artigo 93 (Direitos dos criadores de bancos de dados)

1 O criador de um banco de dados terá o direito de reproduzir, distribuir, difundir ou transmitir (doravante denominado neste artigo como "reprodução") a totalidade ou uma parte substancial de seu banco de dados.

2 Materiais individuais do banco de dados não serão considerados como parte substancial do banco de dados nos termos do parágrafo (1). No entanto, mesmo que a reprodução de materiais individuais ou partes do banco de dados que não constituam uma parte substancial do banco de dados seja repetida ou sistematicamente realizada para uma finalidade específica, ela será considerada uma parte substancial do banco de dados se entrar em conflito com o uso geral do banco de dados ou prejudicar indevidamente os interesses do criador do banco de dados.<Alterado em 18 de maio de 2021, 8 de agosto de 2023>.

3 A proteção prevista neste Capítulo não afetará os direitos autorais sobre o material que faz parte do banco de dados ou qualquer outro direito protegido por esta Lei.

4 A proteção prevista neste capítulo não se estende ao material que faz parte do próprio banco de dados.

(...)

Artigo 94 (Limitação dos direitos dos criadores de bancos de dados)

1 Os artigos 23, 28 a 34, 35(2), 35(4), 35(5), 36 e 37 se aplicarão ao uso de bancos de dados que sejam objeto de direitos dos criadores de bancos de dados. <Alterado em 2 de dezembro de 2011, 26 de novembro de 2019>.

2 Qualquer pessoa pode reproduzir, distribuir, difundir ou transmitir a totalidade ou uma parte substancial do banco de dados em qualquer um dos seguintes casos. No entanto, isso não se aplicará se for inconsistente com o uso geral do banco de dados. <Revisão 18 de maio de 2011, 08 de agosto de 2023>

1. uso para fins educacionais, acadêmicos ou de pesquisa. No entanto, isso não se aplicará se a finalidade for lucrativa.

2. para uso em reportagens sobre assuntos atuais

(...)

Artigo 95 (Período de proteção)

1 Os direitos do criador do banco de dados terão início após a conclusão da criação do banco de dados e continuarão por um período de cinco anos a partir do ano seguinte.

2 Se for feito um investimento humano ou material substancial para atualizar o banco de dados, o direito do criador do banco de dados a essa parte do banco de dados surgirá a partir do momento dessa atualização e durará cinco anos, a partir do ano seguinte.<Revisado em 2021. 5. 18.> (Coreia do Sul, 2023. Tradução feita pelo autor).

1.5. IA Generativa

Um ponto de especial destaque para este trabalho é a menção, ainda que breve, de um tópico relacionado à inteligência artificial e que – conforme será abordado nos capítulos 3 e 5 – tem ocupado espaço proeminente em discussões regulatórias mais recentes no campo do Direito Autoral. Trata-se da IA Generativa – Generative AI (ou GenAI), no inglês.

Resumidamente, pode-se nomear a IA Generativa como qualquer modelo de IA – a exemplo das redes adversariais generativas, vistas anteriormente – cujo resultado envolve a *geração* de diversos tipos de conteúdo em formato de imagens, textos e som (Bridy, 2016; Foster, 2019; Schirru, 2020). Trata-se de uma aplicação da inteligência artificial que comumente utiliza obras protegidas para o seu treinamento – com resultados que parecem desafiar, por vezes, a noção de mineração de textos e dados como uso inerentemente não-expressivo ou não-danoso no âmbito do Direito Autoral.

Exemplos desse tipo de tecnologia já eram descritos na década de 2010 – como indicam os trabalhos de artistas explorando inteligência artificial descritos por Carboni (2011) e o projeto The Next Rembrandt (2016), dedicado a ensinar uma máquina a criar novas pinturas no estilo do célebre pintor holandês de mesmo nome. Contudo, o termo Generative AI parece ter encontrado um uso mais expressivo a partir do final de 2022: em pesquisa no Google Trends, nos últimos cinco anos, foi possível notar que o uso de “Generative AI” como termo de busca experimentou um aumento exponencial de buscas a nível mundial a partir do segundo semestre de 2022 – quando antes o termo encontrava pouca ou nenhuma menção –, com um pico em meados de 2023 e oscilando desde então, porém mantendo-se em níveis altos de busca em comparação com anos anteriores.

FIGURA 5: Gráfico do Google Trends sobre o uso mundial do termo "Generative AI" em buscas entre setembro de 2021 e janeiro de 2025



Fonte: Google Trends. Pesquisa feita pelo autor

A OpenAI é um exemplo de entidade no setor privado que desenvolve diversos produtos oriundos desse tipo de tecnologia. A empresa, em seu site oficial, afirma ser “uma

empresa de investigação e implementação de IA”, cuja missão “é garantir que a inteligência artificial geral beneficie toda a humanidade”.⁵⁹ Em última instância, a companhia se concentra em pesquisas para o desenvolvimento de sistemas de inteligência artificial geral.⁶⁰ Seu portfólio atual inclui ferramentas como o ChatGPT, que utiliza modelos de linguagem⁶¹ para seguir instruções inseridas por humanos – o que inclui sustentar conversas, resumir textos e explicar conceitos. (OpenAI, 2024b). Outro exemplo de destaque é o DALL-E, uma rede neural que utiliza comandos de texto para gerar imagens novas, criadas com base em um corpus de bilhões de imagens associadas a textos (OpenAI, 2024c).

Tais modelos carregam consigo uma questão de particular impacto para o Direito Autoral: embora modelos de IA Generativa como o Next Rembrandt sejam treinados com um acervo de obras curado para apenas incluir material em domínio público, outros sistemas – como os desenvolvidos pela empresa Meta⁶² – por vezes utilizam material que também inclui obras protegidas por direitos autorais. A própria OpenAI reconhece que a GenAI pode trazer impactos econômicos significativos para “certas profissões”.⁶³

Desta forma, o crescimento quantitativo e qualitativo desses modelos tem se demonstrado um fator de preocupação para a classe artística: na indústria musical, uma pesquisa recente conduzida na Alemanha e na França aponta que a maioria dos artistas entrevistados vê a IA como um risco ou ameaça e deseja algum tipo de remuneração que garanta transparência, consentimento e compensação pelo uso de suas obras. O mesmo estudo aponta que, até 2028, cerca de 30% das receitas obtidas nesse setor deverão ser comprometidas pela adoção de sistemas de IA Generativa.⁶⁴ Outro estudo, encomendado pela Authors Guild (uma das maiores entidades de representação de escritores nos EUA), aponta uma preocupação semelhante no campo literário, com a grande maioria dos autores temendo que editoras venham a utilizar

⁵⁹ “OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity” (OpenAI, 2024a).

⁶⁰ “We believe our research will eventually lead to artificial general intelligence, a system that can solve human-level problems. Building safe and beneficial AGI is our mission.” (OpenAI, 2024f).

⁶¹ Ver Ouyang et al. (2022).

⁶² Neste caso em particular, descobriu-se que executivos da Meta, em comunicações internas, demonstraram saber que o processo de treinamento de seus sistemas incluía material não apenas protegido, mas obtido a partir de acervos de conteúdo pirateado – motivando a reabertura de diversos processos contra a empresa (Brittain, 2025).

⁶³ “We recognize that work involving generative models has the potential for significant, broad societal impacts. In the future, we plan to analyze how models like DALL·E relate to societal issues like economic impact on certain work processes and professions, the potential for bias in the model outputs, and the longer term ethical challenges implied by this technology.” (OpenAI, 2024c).

⁶⁴ Os resultados do estudo podem ser visualizados em: <https://www.gema.de/documents/d/guest/gema-sacem-goldmedia-ai-and-music-pdf>. Acesso em 03 out. 2024.

majoritariamente conteúdo gerado por IA (no todo ou em parte), colocando suas carreiras em risco (Authors Guild, 2023b).

Diante deste cenário, diversos setores artísticos já vêm se manifestando de forma contundente: entidades de representação de artistas nos EUA e na Europa já demandaram a diversas empresas de tecnologia pela obtenção de consentimento para o uso de obras protegidas em seus sistemas de IA, bem como uma remuneração justa por estes usos. (Authors Guild, 2023a; European Guild for Artificial Intelligence Regulation, 2023; Initiative Urheberrecht, 2023). Também em 2023, os EUA foram palco de uma greve geral de atores e roteiristas, a qual durou quase cinco meses e tinha, entre suas pautas, demandas para que estúdios se comprometessem a não utilizar conteúdo gerado por inteligência artificial em suas produções (SAG-AFTRA, 2023; Zhuang, 2023).

Além disso, sistemas de IA Generativa já têm sido alvo de disputas judiciais envolvendo agentes de diferentes mídias, como artistas visuais,⁶⁵ fotógrafos,⁶⁶ escritores,⁶⁷ gravadoras musicais⁶⁸ e agências de notícias.⁶⁹ As reclamações incluem a mera reprodução e extração de elementos expressivos no material protegido sem autorização para o treinamento de sistemas de IA,⁷⁰ a subsequente produção de conteúdo que substitui ou compete com as obras originais – seja reproduzindo elementos expressivos das obras ou criando adaptações a partir do estilo artístico de determinados autores – e a remoção de conteúdo que identifique os titulares.⁷¹ Nos Estados Unidos, uma versão compilada dos diversos litígios em curso no país até dezembro de 2024 pode ser vista em Knibbs (2024).

A maior parte dessas manifestações parece se embasar, em primeiro lugar, no efeito substitutivo que a IA Generativa tem exercido sobre os artistas: a geração de conteúdo de forma automatizada, afinal, precluiria a requisição dos serviços de um profissional. Ao mesmo tempo,

⁶⁵ Por exemplo, ver: Andersen v. Stability AI, U.S. District Court for the Northern District of California, No. 3:23-cv-00201. Disponível em: <https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.1.0.pdf>. Acesso em 29 fev. 2024.

⁶⁶ Por exemplo, ver Estados Unidos (2023b).

⁶⁷ Por exemplo, ver: Estados Unidos (2023e; 2024e).

⁶⁸ Vide, por exemplo, Estados Unidos (2024d; 2024h).

⁶⁹ Ver Estados Unidos (2023f; 2024f; 2024g).

⁷⁰ Exemplo no caso Silverman: “Much of the material in OpenAI’s training datasets, however, comes from copyrighted works—including books written by Plaintiffs—that were copied by OpenAI without consent, without credit, and without compensation” (Estados Unidos, 2023e, p. 5).

⁷¹ Ver, por exemplo, no caso Raw Story v. Altnet Media: “When they populated their training sets with works of journalism, Defendants had a choice: they could train ChatGPT using works of journalism with the copyright management information protected by the DMCA intact, or they could strip it away. Defendants chose the latter, and in the process, trained ChatGPT not to acknowledge or respect copyright, not to notify ChatGPT users when the responses they received were protected by journalists’ copyrights, and not to provide attribution when using the works of human journalists.” (Estados Unidos, 2024f, p. 2).

contudo, esse efeito substitutivo só seria observável em razão do grau de sofisticação hoje alcançado por diversos sistemas de IA – e este, por sua vez, só seria possível graças ao uso das obras desses mesmos artistas no processo de treinamento (Geiger, 2022; Senftleben, 2023). Logo, a mobilização do Direito Autoral se justificaria como forma de garantir o atendimento de direitos fundamentais dos artistas, os quais estariam comprometidos pelo uso da IA Generativa (Geiger e Iaia, 2024). Tratam-se de argumentos lógicos, e que fazem sentido quando consideramos que o efeito substitutivo da IA reside para além da substituição da obra, foco da proteção do Direito Autoral: aqui, o próprio autor é suplantado.

Entretanto, talvez por esse mesmo motivo, é pertinente adiantar que a demonstração factual de uma violação em um caso específico (como nos litígios vistos acima) pode representar um processo deveras complexo, especialmente no tocante a considerações relativas à originalidade e à distinção entre ideia e expressão. Guadamuz (2024, p. 8; 19-25), por exemplo, se manifesta novamente ao reconhecer que o processo de treinamento de inteligências artificiais pode envolver atos que interfiram com direitos exclusivos, mas que há duas dificuldades em comprová-lo na prática: em relação ao treinamento, é difícil identificar o efeito de uma obra em particular sobre um modelo de IA, já que estes utilizam coleções imensas e os resultados dependem muito mais da informação agregada do que de dados específicos. Do ponto de vista das obras produzidas por GenAI, é possível observar resultados similares a obras originais que indiquem a capacidade de os modelos memorizarem certos elementos expressivos das obras e de serem ensinados a esquecê-los – geralmente no caso de obras ou códigos populares, que possuem muitas variantes capturadas em treinamento.⁷² Mas a incidência de reproduções ou adaptações claras tem se mostrado baixa,⁷³ e ainda não há mecanismos confiáveis para identificar a relação causal entre uma obra específica e um output de IA em particular.

Isto não quer dizer que não existam tentativas no sentido de prover ao menos alguma forma mais ampla de informação para os autores. Por exemplo, o site “Have I Been Trained?”⁷⁴ permite que os artistas verifiquem se um trabalho específico foi incluído no LAION-2B – uma versão mais antiga do LAION-5B – e permite que os artistas solicitem a remoção do trabalho do banco de dados. Ainda assim, trata-se de uma iniciativa com limitações evidentes: em primeiro lugar, a ferramenta não permite buscas em massa: é necessário pesquisar pelos metadados de uma obra por vez, o que torna o processo de pesquisa mais demorado. Em segundo lugar,

⁷² Para um exemplo, ver Eldan e Russinovich (2023).

⁷³ Vide Somepalli G. et al., 2022; Carlini N. et al., 2023.

⁷⁴ Have I Been Trained? (2024).

apontar a existência de metadados em uma base não implica em ser capaz de indicar como eles foram usados. Em outras palavras, se um sistema como o Midjourney faz uso desta base, o “Have I Been Trained?” ainda não informa de que maneira exata uma imagem gerada por IA utilizou algum conteúdo em particular.

Diante destes problemas técnicos, nota-se que as cortes de origem dos litígios descritos acima têm mostrado reticência em relação a alguns desses argumentos, especificamente no que diz respeito ao que, exatamente, configuraria violação nos outputs de IA Generativa – o que tem motivado pedidos de emenda em petições. No caso *Silverman*, por exemplo, entendeu-se o uso das obras não implicou em produtos finais que possuíssem similaridades suficientes com o original a ponto de substituí-los (Brittain, 2023).⁷⁵ Já no caso *Andersen*, parte substancial dos comentários do juízo se concentrou sobre a dificuldade em provar que as imagens geradas pelos programas desenvolvidos pelos réus fizeram uso substancial da expressão contida nas obras usadas, ao invés de tão somente o “estilo artístico” dos autores.⁷⁶

Por outro lado, dois casos mais recentes apontam para esforços nas indústrias criativas no sentido de contornar esta questão: em um deles, o grupo editorial HarperCollins teria entrado em um acordo com um desenvolvedor não-mencionado no sentido de licenciar parte de seu catálogo para fins de treinamento de IA por um período de três anos – com a condição de que os autores das obras envolvidas deveriam autorizar previamente a inclusão das mesmas (Peters,

⁷⁵ Ver Brittain, B. (2023).

⁷⁶ Decisão preliminar do caso *Andersen*, p. 13-13: “Even if that clarity is provided and even if plaintiffs narrow their allegations to limit them to Output Images that draw upon Training Images based upon copyrighted images, I am not convinced that copyright claims based a derivative theory can survive absent “substantial similarity” type allegations. The cases plaintiffs rely on appear to recognize that the alleged infringer’s derivative work must still bear some similarity to the original work or contain the protected elements of the original work. See, e.g., *Jarvis v. K2 Inc.*, 486 F.3d 526, 532 (9th Cir. 2007) (finding works were derivative where plaintiff “delivered the images to K2 in one form, and they were subsequently used in the collage ads in a quite different (though still recognizable) form. The ads did not simply compile or collect Jarvis’ images but rather altered them in various ways and fused them with other images and artistic elements into new works that were based on- i.e., derivative of-Jarvis’ original images.”) (emphasis added); *ITC Textile Ltd. v. Wal-Mart Stores Inc.*, No. CV122650JFWAJWX, 2015 WL 12712311, at *5 (C.D. Cal. Dec. 16, 2015) (“Accordingly, even if Defendants did modify them slightly, such modifications are not sufficient to avoid infringement in a direct copying case.... Thus, the law is clear that in cases of direct copying, the fact that the final result of defendant’s work differs from plaintiff’s work is not exonerating.”) (emphasis added); see also *Litchfield v. Spielberg*, 736 F.2d 1352, 1357 (9th Cir. 1984) (“a work is not derivative unless it has been substantially copied from the prior work”); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 225 (2d Cir. 2015) (“derivative works over which the author of the original enjoys exclusive rights ordinarily are those that re-present the protected aspects of the original work, i.e., its expressive content”). Defendants make a strong case that I should dismiss the derivative work theory without leave to amend because plaintiffs cannot plausibly allege the Output Images are substantially similar or re-present protected aspects of copyrighted Training Images, especially in light of plaintiffs’ admission that Output Images are unlikely to look like the Training Images. Compl. ¶ 93. But other parts of plaintiffs’ Complaint allege that Output Images can be so similar to plaintiff’s styles or artistic identities to be misconstrued as “fakes.” Compl. ¶¶ 44ii, 171-172. Once plaintiffs amend, hopefully providing clarified theories and plausible facts, this argument may be re-raised on a subsequent motion to dismiss.” (Estados Unidos, 2023d).

2024). Já na esfera judicial, a Universal Music Group entrou em acordo com a desenvolvedora Anthropic, pelo qual a ré passaria a manter ao seu sistema de IA “travas” tecnológicas que impeçam o programa de prover letras de músicas de propriedade das gravadoras envolvidas no litígio (Cho, 2025).

Outro litígio mais antigo, de 2020, também pode gerar precedentes de interesse para a mineração de textos e dados nos Estados Unidos: o caso Thomson-Reuters v. Ross Intelligence (Estados Unidos, 2020), que também tem no seu cerne a inteligência artificial, porém com caráter não-generativo. Aqui, os proprietários do serviço de pesquisa jurídica Westlaw acusam a empresa Ross Intelligence, de utilizar uma empresa terceirizada) para copiar o conteúdo do repositório da Westlaw – composto, em grande maioria, por documentos jurídicos, que o Direito Autoral dos EUA, como no Brasil, exclui de proteção por Direito Autoral – para treinar um sistema de inteligência artificial concorrente. O objeto da ação também inclui as anotações feitas aos autos de processos, bem como a organização e classificação do conteúdo.⁷⁷ Trata-se, portanto, de um caso que não apenas exemplifica como o uso de obras protegidas em IA (e, por extensão, em mineração de textos e dados) pode ser alvo de disputas judiciais, como também ilustra de que maneira o Direito Autoral sobre bases de dados pode servir para obstaculizar o uso de conteúdo que, em teoria, se encontraria fora da seara desse instituto.

Entretanto, é a China que aparenta haver trazido uma resposta preliminar para este caso, através de uma decisão de fevereiro de 2024 emitida pela Corte de Internet de Guangzhou. O caso em questão envolve a responsabilização judicial de uma empresa de inteligência artificial em razão de seu modelo haver gerado imagens similares às do super-herói japonês Ultraman, pelo qual decidiu-se que a IA teria infringido os direitos exclusivos de reprodução da obra pelo titular ao gerar imagens substancialmente similares às do material original, determinando que o software seja tornado incapaz de gerar resultados similares mesmo quando acionado nesse sentido (Costigan, 2024). Note-se que, diferente dos casos acima, a decisão parece ter considerado não o treinamento em si como infringente, mas unicamente o output gerado pelo modelo de IA.

⁷⁷ “Integral to Westlaw is Plaintiffs’ WKNS [West Key Number System], which organizes U.S. law using a hierarchy that is unique to Plaintiffs. The WKNS is the backbone through which thousands of lawyers conduct legal research. The development of the WKNS, beginning in print and now in a digital format, has been and continues to be the result of Plaintiffs’ numerous creative choices about how to organize cases and which cases to place in that classification, requiring substantial investments of time, technological and human resources, and money over the course of decades.” (Estados Unidos, 2020).

Já no Brasil, ao menos parte desta discussão encontra precedente jurisprudencial: o Superior Tribunal de Justiça – que, como visto anteriormente, rejeitou a noção de que informações científicas constituiriam material protegido – se posiciona de forma bem explícita ao considerar que estilos e técnicas artísticas não são passíveis de proteção pelo Direito Autoral, equiparando-as a ideias, métodos e procedimentos e, portanto, tratando tais entes como distintos da expressão artística resultante dos referidos processos⁷⁸ Contudo, destaca-se que tal discussão tem impacto majoritariamente sobre o produto final (o *output*) da coleta e análise de dados realizada por sistemas de IA. Ainda não há menção, nas cortes superiores brasileiras, sobre como a coleta, armazenamento e análise em si mesmos (o *input*) se situam no Direito Autoral pátrio.

Nesse sentido, é importante lembrar que, no Brasil, é justamente a legalidade do *input* em si que é posta em xeque: conforme visto anteriormente, o texto do art. 29, X da Lei de Direito Autoral é categórico ao afirmar que a mera inclusão de obras protegidas em bases de dados necessita de prévia e expressa autorização do titular (Brasil, 1998b). Assim, a condução de processos-padrão de mineração de textos e dados na pesquisa dentro e fora do território de IAs avançadas dependerá de uma futura regulação, legal ou judicial, que estipule as circunstâncias que permitam o uso de obras protegidas para esses fins.

Nesse sentido, os próximos dois capítulos se debruçam especificamente sobre como o Direito Autoral pode tratar deste tipo de situação, considerando a disponibilidade de instrumentos legais de garantia ao interesse público e a outros direitos fundamentais, como no caso das limitações e exceções aos direitos de autor e conexos. Veremos, assim, as possibilidades oferecidas por este instituto diante do arcabouço jurídico em vigor, sua aplicação ao redor do mundo e de que forma o atual cenário tecnológico e de políticas públicas afeta a sua discussão e implementação.

⁷⁸ “RECURSO ESPECIAL. OFENSA AOS ARTS. 458 E 535, I E II, DO CPC. INEXISTÊNCIA. LITIGÂNCIA DE MÁ-FÉ. PENALIDADES MAL APLICADAS. DIREITOS AUTORAIS. ESTILOS, MÉTODOS OU TÉCNICAS. INEXISTÊNCIA DE PROTEÇÃO.

1. Não ofende o Art. 535 do CPC o acórdão que, embora rejeitando os embargos de declaração, examinou todas as questões pertinentes.

2. Não é nula, por falta de fundamentação, sentença na qual o juiz declina completamente os motivos de seu convencimento.

3. A aplicação de penalidades por litigância de má-fé exige dolo específico, perfeitamente identificável a olhos desarmados, sem o qual se pune indevidamente a parte que se vale de direitos constitucionalmente protegidos (ação e defesa).

4. **Estilos, métodos ou técnicas não são objetos de proteção intelectual (Art. 8º, I e II, da Lei 9.610/98). O que se tem sob guarida legal são as obras resultantes da utilização de estilos, métodos ou técnicas.**” (Brasil, 2007, p. 228. Destaque nosso).

CAPÍTULO 2: LIMITAÇÕES E EXCEÇÕES PARA MINERAÇÃO DE TEXTOS E DADOS NA ESFERA SUPRANACIONAL

No Capítulo 1, observou-se de que forma a mineração de textos e dados adentra o território do Direito Autoral, especialmente à luz dos desenvolvimentos recentes em inteligência artificial. Agora, voltamo-nos para como o uso de obras protegidas por Direito Autoral para fins de pesquisa vem sendo abordado nos debates internacionais e regulações sobre Direito Autoral ao redor do mundo.

Para tanto, dividimos esta análise em duas partes: neste Capítulo, debruçamo-nos especificamente sobre as bases da propriedade intelectual enquanto regime de apropriação de bens imateriais, bem como o papel do regime de limitações e exceções de Direito Autoral (L&E) nesse regime de uma forma mais ampla – uma vez que vêm sendo uma forma amplamente utilizada ao redor do mundo para regular a mineração de textos e dados para fins de pesquisa. Por fim, trazemos algumas considerações mais críticas a respeito da forma como as limitações e exceções vêm sendo mobilizadas, chamando atenção para a forma como isto impacta a forma de se implementar permissões para uso de obras na pesquisa brasileira.

Desta forma, este Capítulo se concentra em discussões de cunho mais amplo, na arena supranacional: as variadas maneiras pelas quais diferentes regiões no mundo vêm abordando o assunto, bem como discussões sobre o tema a nível nacional e regional, serão abordadas no Capítulo seguinte, quando analisaremos os diferentes modelos de regulamentação da L&E para mineração de textos e dados ao redor do mundo.

2.1. Conhecimento, regimes de apropriação e Direito Autoral

Compreender o uso das limitações e exceções ao Direito Autoral como ferramenta de incentivo à pesquisa e à inovação requer, em um nível mais básico, uma compreensão acerca de como diferentes visões sobre apropriação de bens se desenvolveram ao longo do tempo e como isto influencia a gestão do conhecimento. Para tanto, esta seção inicia suas observações a partir das mudanças na concepção de propriedade a partir da Idade Contemporânea e como isto influencia os contornos assumidos pelo Direito Autoral ao longo do tempo.

Assim, esta análise parte da ideia de que o direito de propriedade nem sempre existiu na forma como o conhecemos atualmente: Hespanha observa que, durante o Séc. XIV, predominava a noção de que tudo e todos estavam sujeitos a uma ordem universal, na qual cada elemento cumpria um papel específico um “*summum bonum*”. Desta forma, cada coisa possuía diferentes usos e cada pessoa ligada a essa coisa recebia, portanto, prerrogativas diferentes de

acordo com sua função específica: aos nobres, cabia proteger a terra e receber tributos; aos camponeses, cabia ali morar, produzir e pagar esses tributos (Hespanha, 1980, p. 219-220).

Entretanto, na mesma época, cresciam correntes filosóficas centradas em uma concepção mais individualista e antropocêntrica, situando o ser humano como um ser dotado de liberdade – esta, por sua vez, manifestada nas coisas que são ligadas a esse homem. Ora, por este pensamento, o homem só poderia ser pleno se possuísse o controle total, livre de intervenções externas e, principalmente, concentrado apenas sobre este indivíduo. Assim, a propriedade deveria ser a) um direito natural e absoluto, ligado à própria existência do ser humano ao invés de ser condicionado a uma ordem externa; b) pleno, podendo ser exercido de forma total, em todos os seus usos possíveis, mesmo os não-econômicos; d) em teoria, perpétuo, para sempre (ou por muito tempo) ligado àquele indivíduo de alguma forma e e) privado em essência, sem submissão à coisa pública. (*Ibid.*, p. 211-212, 219).

Tal pensamento viria a crescer com a ascensão do pensamento liberal no final do século XVIII, aparecendo em filósofos como Locke (1764) – para quem o homem, quando exerce labor sobre algo que advém do domínio comum (ex.: colher uma fruta de uma árvore na selva), cria uma ligação com aquela coisa, tornando-a sua (Locke, 1764) – e Kant, para quem a posse sobre uma coisa está ligada ao respeito à liberdade individual, que é universal e absoluta, não aceitando relativizações (Kant, 2013, p. 38-39). Tal pensamento viria a se cristalizar juridicamente com os movimentos de codificação da vida social no séc. XIX, através da consolidação, nos códigos civis europeus, de um sistema “monocultural” de propriedade como poderes exclusivos e dispositivos – interpretando como anomalia a persistência de modelos coletivos de gestão de bens não apenas fora do eixo europeu, mas mesmo dentro dele (Grossi, 2006, p. 8-9)⁷⁹.

Associada a uma noção de história como uma sucessão de estágios que tende a um “progresso”, esta visão linear e unidirecional de um modelo privatista de apropriação como o ideal e “natural” ganhou espaço cada vez maior no âmbito cultural, econômico e social

⁷⁹ “Essa assim chamada ‘propriedade coletiva’ é uma propriedade? Temos certeza de usar legitimamente um tal termo para a sua descrição? Ou estamos diante de um esquema ordenador que parte de premissas opostas e, portanto, irreduzíveis também em uma propriedade relativizada no mais extenso dos plurais? A velha civilística do séc. XIX, elevada em seu êxtase individualista, contemplou-a com a mesma suspeita curiosidade com a qual um naturalista analisa um monstinho sobre sua mesa de experiências, sigilou-a como uma ‘anomalia’, e dela sentiu repugnância; não repugnância mas uma individualista intolerância continuou a circular na civilística do séc. XIX, sobre um perfil interpretativo que de Bonelli chega diretamente a Salvatore Romano, onde, sem se preocupar com uma possível inserção em outros universos culturais, afirma-se que a assim chamada propriedade coletiva não realiza um esquema proprietário somente pela incapacidade desses juristas de admitir o aviltamento coletivo do poder dispositivo” (Grossi, 2006, p. 8-9).

(Congost, Gelman e Santos, 2012, p. 7), de forma que sistemas de apropriação baseados em uma gestão descentralizada passaram a ser vistos como vulneráveis ao parasitismo e dependentes de elevados custos de transação (Ostrom e Hess, 2007). Um exemplo pode ser visto na “tragédia dos comuns” popularizada por Hardin: para o autor, bens de uso coletivo estariam naturalmente vulneráveis a ações predatórias por parte dos usuários, uma vez que, em uma percepção individual, os benefícios do uso excessivo são mais visíveis do que os malefícios (que são diluídos dentro do grupo como um todo). Assim, cada membro seria tentado a extrair um pouco mais do que caberia a si, culminando na degradação gradual do recurso em questão – a menos que houvesse uma limitação externa sobre o seu uso (Hardin, 1968, p. 1244-1245).

Este modelo, contudo, encontrou crescente oposição, sobretudo por movimentos no ambiente jurídico e econômico favoráveis a uma gestão mais fragmentada de direitos (Congost, Gelman e Santos, 2012, p. 7). Nesse sentido, Ostrom (1990) já criticava o aparente aprisionamento teórico experimentado por teóricos e gestores que limitava a administração de bens de uso coletivo a um sistema binário: ora entendendo que a tragédia dos comuns só poderia ser evitada se esses recursos fossem geridos pelo Estado ao invés de serem deixados nas mãos de indivíduos; ora apontando o Estado como um gestor limitado em termos de informação, capacidade de monitoramento e *enforcement*. Ostrom propõe, portanto, que não há uma “única forma possível” de gerir bens: há diversas possibilidades que funcionarão melhor ou pior dependendo do caso (Ostrom, 1990, p. 14). Ademais, duas décadas depois, a autora ainda observaria, por meio de observações empíricas, que é possível ter sistemas *mistos*, onde diferentes regimes de apropriação podem coexistir em níveis diferentes dentro de um mesmo ambiente (como em uma cooperativa de agricultores onde cada um tem seu terreno, mas todos são coletivamente beneficiários e responsáveis por um sistema de irrigação comunitário, que os abastece) (Ostrom, 2010).

Nesse sentido, Ostrom e Hess (2007) defenderiam a propriedade não como um direito unificado, mas enquanto um “feixe de direitos”, tradicionalmente identificados em cinco: acesso, extração (de recursos extraídos daquele bem), gerenciamento, exclusão e alienação (a concessão do bem, no todo ou em parte, para outrem). Para as autoras, estes direitos podem ser distribuídos e compartilhados dentro de um grupo de diversas maneiras, chamadas de *regimes de* apropriação. Estes incluem modalidades nas quais um bem subsiste melhor e é mais frutífero quando gerido por um único ente, privado ou estatal, e também regimes “no meio do caminho” no qual os direitos sobre um dado bem parecem funcionar adequadamente quando amplamente distribuídos e geridos internamente, por um grupo de indivíduos ou entidades, que

administrariam o bem de forma coletiva. As autoras (*Ibid.*, p. 6) também ressaltam que a gestão coletiva não equivale a regimes de acesso irrestrito – este meramente diz respeito à inexistência de regras de exclusão ao acesso, enquanto um bem gerido coletivamente pode ter seu acesso e uso controlado pelos membros.

A necessidade de se considerar uma maior diversidade de modelos de gestão da propriedade se faz mais relevante quando se levam em conta as considerações de Ostrom (2010, p. 645) no sentido de que os próprios bens a serem geridos por esses modelos possuem naturezas diversas em relação à sua rivalidade (o quanto o valor de um bem é consumido quando usado por alguém) e de capacidade de exclusão (o quão fácil é restringir seu acesso e uso por terceiros). Dessa forma, a autora encontra quatro categorias de bens: os estritamente privados, como alimentos e roupas, de alta rivalidade (o uso por um indivíduo impede o uso por outrem) e cujo uso pode ser restrito facilmente, os recursos de uso comum (ou *common-pool resources*), como reservas ambientais (que se exaurem fácil, mas têm acesso difícil de controlar), os “bens de clube”, como teatros (cujo uso pode ser compartilhado entre múltiplos indivíduos, mas podem ter beneficiários mais facilmente controlados) e, por fim, os bens públicos, assim chamados porque são de difícil exclusão e seu uso pode ser compartilhado por uma quantidade indefinida de indivíduos (não confundir com os bens públicos no sentido jurídico).

Aqui, é importante destacar que esta classificação trata exclusivamente da natureza dos bens, e não dos regimes de apropriação dos mesmos: “bens públicos” e “bens privados”, aqui, devem ser entendidos no seu sentido econômico, referente à natureza coletiva ou individual do seu uso, diferindo do sentido jurídico do termo (que se refere à gestão de um bem pelo Estado ou centralizada em um indivíduo). Ainda assim, esta diferenciação – e a presença de bens intermediários, como os bens de clube e de uso comum – serve para reforçar a inadequação de se reduzir as possibilidades de gestão de bens a um ou outro modo de apropriação.

Dentro desta classificação, nota-se que o conhecimento aparece na literatura como um bem público (Coriat, 2013; Abecassis et al, 2019). Mais especificamente, Coriat (2013, p. 17-19) expande as observações de Ostrom em relação a este tipo de recurso, referindo-se a acervos de dados, obras e outros bens informacionais como *knowledge commons*. Para o autor, estes bens se comportam de forma bastante distinta dos *commons* tradicionais, de natureza física: eles não teriam simplesmente uma rivalidade baixa, mas seriam completamente não-rivais. Conhecimento não se reduz em valor ou em existência quando alguém usufrui do mesmo; porém, mais do que isso, quanto mais o conhecimento é adquirido e usado, mais ele se replica e expande. A gestão dos *knowledge commons* passa a ser pautada não à conservação daquele

patrimônio, mas ao seu enriquecimento e crescimento – de forma que Ostrom e Hess (2007, p. 16) identificam, nesses tipos de recurso, mais dois direitos além dos cinco usualmente vistos em outros tipos de bens: o de contribuir (com o material de um repositório) e o de removê-lo.

Estas diferenças fundamentais na natureza dos *knowledge commons* também implicam em dizer que o conhecimento, por natureza, não parece se conformar ao tipo tradicional de recurso que estaria sujeito à tragédia dos comuns – e que, pelo paradigma liberal, requereria uma concentração e restrição dos direitos de acesso e uso. Ao contrário, estudos da década passada alertaram que os bens informacionais têm corrido um risco específico de degradação, em um fenômeno chamado de “tragédia dos *anticommons*”, pelo qual a expansão excessiva de direitos de propriedade intelectual – direito de autor e patenteamento predatório inclusos – tem levado a um “engessamento” da inovação ao limitarem o compartilhamento de informação (Nelson, 2003, p. 18, 22-25; Ostrom e Hess, 2007b, p. 11; Hess, 2008, p. 7; Coriat, 2013, p. 15). Mais especificamente, seguindo uma lógica análoga à de Hardin, se os benefícios individuais à concentração de direitos parecem mais palpáveis que os benefícios coletivos oriundos da responsabilidade de se manter um grau de abertura para o acesso e uso do conhecimento pelo público, então haveria uma tendência à degradação desse bem – mas, diferentemente da tragédia dos comuns, esta ocorreria pelo *excesso* de controle, não pela falta dele.

No campo da pesquisa científica, Murray e Stern também observaram este efeito *anti-commons* ao comparar artigos diretamente relacionados a patentes ou a áreas de pesquisa aplicada com trabalhos de áreas mais básicas, onde a ligação com patentes seria menor. Os autores notaram que a quantidade de citações de trabalhos no primeiro grupo ao longo do tempo apresentava uma queda gradual em relação ao segundo grupo, o que os autores atribuíram a restrições trazidas pela proteção de informações patenteadas ou da tecnologia envolvida nas mesmas (Murray e Stern, 2007, p. 683-684).

Nesse sentido, o Direito Autoral, enquanto parte da propriedade intelectual, não escapou destas influências: pelo contrário, estas sempre se manifestaram de diversas formas ao longo da história do instituto. Woodmansee (1984), por exemplo, já abordava o processo de surgimento da figura do autor que acompanhou a Idade Contemporânea e a transformação de sua posição. Se, antes, o autor era visto simplesmente como um artesão habilidoso, um copista cujo valor consistia na sua capacidade de reproduzir a natureza ou, mais tarde, uma inspiração externa (seja Deus ou mesmo uma “musa”), diversos textos de escritores no Séc. XVIII demonstram uma emergência do autor “gênio”: aquele que por uma inspiração ou disposição vinda de si mesmo, cria algo singular a partir do nada. Isto encontraria paralelo com

movimentos na classe artística em países como Alemanha, Inglaterra e França, acompanhando um processo de transformação no mercado de arte, sobretudo de livros: a difusão de obras facilitada pela prensa de Gutenberg, somada ao crescimento de uma classe média com recursos financeiros, fez com que artistas dependessem menos do mecenato e se atraíssem mais pela venda de cópias de suas obras ao grande público. Assim, para os escritores, tornou-se vital estabelecer-se como uma categoria profissional, que prestava um serviço (ou ofertava um produto) à sociedade e merecia ser apropriadamente remunerada em troca.

Já Jaszi (1991) aborda esta questão por um ângulo diferente, afirmando que o instituto do copyright algo-saxônico, já em seus primórdios, trazia a noção de “autoria” e “obra” menos como demandas de artistas e mais como meros fundamentos retóricos mobilizados por livreiros para legitimar a necessidade de garantir um direito exclusivo sobre as vendas de obras. Na verdade, segundo o autor, o que se almejava era criar um instituto a ser negociado com os autores, para então obter o monopólio sobre as vendas de obras, como retorno dos riscos do investimento em imprimi-las e divulgá-las. Ao mesmo tempo, Jaszi ressalta que tais fundamentos retóricos teriam a vantagem de ser facilmente ampliados e tornados cada vez mais abstratos: cada vez mais objetos, como fotografias, passaram a ser incluídos como “obra” e cada vez mais entes, como empresas, passariam a ser “autores” – às vezes, em detrimento das próprias pessoas físicas responsáveis pela criação de uma peça de arte. Os próprios “direitos morais”, como a garantia da atribuição do artista e de integridade da obra, serviriam apenas como um “afago” à classe artística, enquanto editores e outros intermediários permaneceriam os maiores beneficiários econômicos desse direito de propriedade.

Não surpreende, portanto, o movimento de expansão da propriedade intelectual ao final do século passado – o mesmo que incluiu bancos de dados e programas de computador sob a égide do direito de autor, conforme visto anteriormente. Com a expansão das tecnologias da informação e do processo de globalização da economia, diversos autores apontam que os tratados internacionais e regulações postas em vigor na década de 1990 são fruto de movimentos predominantemente de países do Norte Global, visando proteger suas crescentes indústrias intensivas em conhecimento contra concorrentes (Sell, 2003; Orsi e Coriat, 2006; Granstrand, 2007). O resultado teria sido um processo de redução do domínio público, de dificuldade à obtenção de informação e a acentuação de assimetrias e de relações quase neocoloniais entre Norte e Sul Global, com o uso da propriedade intelectual como mecanismo de barreira à participação de países em desenvolvimento naquele momento do processo de inovação global (Sell, 2003; Okediji, 2009; Rahmatian, 2009).

Nesse sentido, Ascensão adota um tom similar ao de Jaszi, mas lamentando o que considera um esvaziamento dos direitos morais, em favor do resguardo aos investimentos de grandes titulares. Para o autor, as verdadeiras razões para a proteção não são de ordem filosófica ou conceitual, mas prática: aproveitar-se da menor burocracia (dada a inexigibilidade de registro das obras) e dos tempos maiores de duração da proteção em relação a outras modalidades de propriedade intelectual. O resultado seria um esgarçamento das concepções de “autor” e “obra” ao ponto de elas se tornarem explicitamente irrelevantes, inexistentes. Se, para Jaszi, o artista nunca realmente importou, Ascensão entendia que o autor agora era abertamente descartado em favor de questões de mercado e do investimento dos grandes titulares (Ascensão, 2008, p. 8-16).

Ao mesmo tempo, essa mesma reação expansionista do Direito Autoral às tecnologias de informação na virada do século se manifestava sob a forma de um combate intenso à pirataria digital, então vista como uma ameaça aos interesses de artistas e titulares. Para além da imposição de medidas anti-burla (vistas no Capítulo 1), esta época também testemunhou grandes litígios sobre serviços de compartilhamento de arquivos entre pares e regulações anti-pirataria – o que, para autores como Carrier (2012) e Frosio (2023), teria culminado em um ambiente resistente à inovação, onde grandes detentores de direitos optavam por um “modelo de negócio” centrado em lucrar com processos e proteger mercados já estabelecidos ao invés de buscar novos modelos de licenciamento e exploração econômica em benefício de artistas e outros titulares.

Por outro lado, a história da propriedade intelectual também serve de evidência de que a oposição a um modelo monolítico de propriedade não é recente: mesmo entre liberais, diversos movimentos em prol do livre comércio ao longo dos séculos XVIII e XIX se posicionavam no sentido de declarar a propriedade intelectual como um obstáculo à livre iniciativa, gerando protestos contra a expansão de legislações de PI que acabaram, todavia, cedendo a pressões políticas e econômicas (Granstrand, 2007, p. 271).

Na literatura científica, há posicionamentos divididos sobre o efeito da apropriação via propriedade intelectual. Uns apontam para uma relação direta e positiva entre direitos de PI e crescimento econômico no longo prazo, indicando que a proteção do patrimônio imaterial traria mais segurança ao trabalho de entidades e profissionais ligados à produção científica e tecnológica (Kwan e Lai, 2003; Eicher e Penalosa, 2008; Arnold e Bauer, 2009). Por outro lado, há autores que, embora reconheçam esta relação positiva, observam que ela possui limites: a partir de um ponto, restrições geradas pela PI ultrapassariam um ponto ótimo, levando a um crescente

monopólio na produção e disseminação de conhecimento e “engessando” a capacidade de alcançar novas descobertas (Furukawa, 2007; Murray e Stern, 2007; Gangopadhyay e Mondal, 2012).

Flynn e Palmedo, por sua vez, empreenderam uma análise quantitativa diferente, medindo a relação entre o grau de abertura de um regime nacional de PI e o grau de desenvolvimento em setores intensivos em conhecimento. Diferentemente de outros autores, o texto aponta para uma relação direta e positiva entre um maior grau de abertura e receitas mais altas nas indústrias da informação, sem representar danos significativos a segmentos que lidam mais diretamente com o Direito Autoral. No campo da produção acadêmica, a mesma correlação positiva foi observada entre direitos autorais mais permissivos e uma produção mais numerosa e com maior grau de citações (Flynn e Palmedo, 2017, p. 16-17; 19-20).

Em outras palavras, o que se observa é que, como sugere Lewicki (2007), o Direito Autoral, embora se coloque como direito natural (e, portanto, atemporal), está necessariamente ligado a um contexto social, político e econômico. Constatase, assim, que o instituto surgiu e tem se desenvolvido primariamente sob os auspícios de um sistema de pensamento que privilegia a privatização e concentração de direitos como forma ideal de gestão da propriedade – o que se mostra problemático quando se considera que o bem a ser gerido neste caso é o próprio conhecimento: um bem imaterial que tem, em última análise, a coletividade como principal beneficiário.

Por esta razão, parece razoável entender que o Direito Autoral opera mais eficientemente não como um direito exclusivamente privatista, mas sim como um regime misto (com partes individualistas e outras coletivas) de organização dos direitos de propriedade, a fim de preservar o estímulo individual à produção de conhecimento e, ao mesmo tempo, estabelecer garantias que coíbam e corrijam ineficiências resultantes do abuso na gestão desse direito e permitam que os frutos dessa produção possam ser mais eficazmente difundidos à sociedade. Dentre os mecanismos empregados para tanto, destacam-se as limitações e exceções ao Direito Autoral.

2.2. Uma breve introdução sobre limitações e exceções no Direito Autoral

O Direito Autoral, enquanto direito de propriedade, se manifesta sob a forma de uma exclusividade temporária sobre a exploração de uma obra artística e literária, visando garantir a artistas e titulares a segurança de poder usufruir economicamente (e, portanto, subsistir) da

sua produção (Chaves, 1987, p. 20), ao mesmo tempo em que há uma preocupação com o resguardo da autoria e a integridade da obra (no âmbito dos direitos morais).⁸⁰

Contudo, nenhuma forma de propriedade é absoluta – tampouco o Direito Autoral. Mesmo nos debates públicos que levaram à construção do instituto na idade moderna, era possível ver os direitos exclusivos dos autores sendo situado por comentaristas como uma exceção à regra natural de que o conhecimento era livre. Woodmansee (1984, p. 438) menciona como exemplo a separação entre ideia e expressão: críticos da época já afirmavam que, mesmo que a obra seja distinta do livro que a materializa, isto não implica no controle sobre as ideias contidas em um livro, mas apenas sobre a *forma* específica assumida por elas – do contrário, o próprio leitor não poderia adquirir um livro para adquirir o conhecimento ali contido.

Já uma literatura mais contemporânea situa este debate nos termos de uma ordem constitucional pautada no atendimento ao interesse público, à qual o Direito Autoral pertence e deve deferência. Nesse sentido, Souza, A. (2005) aplica o princípio constitucional da função social da propriedade sobre os direitos de autor e conexos, colocando-os a serviço dos interesses da coletividade – interesses estes principalmente relacionados ao direito à educação, informação e cultura, mas que podem ser ampliados para outros campos, como direito do consumidor, livre concorrência e soberania nacional. A proteção garantida pelo Direito Autoral deve, portanto, se dar na medida em que o alcance destes interesses seja respeitado. Do contrário, conforme Souza, C. (2009, p. 204-277) observa-se a incidência do abuso do Direito Autoral – que parte do conceito de abuso de direito existente no Código Civil, quando ocorre um desvio de finalidade no exercício de um dado direito, violando sua função social, a boa-fé ou dos bons costumes. Desta forma, o abuso do Direito Autoral se coloca como ferramenta jurídica a ser mobilizada pelo Judiciário em caso de emprego abusivo dos direitos de exclusividade a fim de restituir a conformidade dos direitos de autor com esta função social.

⁸⁰ “Article 6bis

[Moral Rights: 1. To claim authorship; to object to certain modifications and other derogatory actions; 2. After the author’s death; 3. Means of redress]

(1) Independently of the author’s economic rights, and even after the transfer of the said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation.

(2) The rights granted to the author in accordance with the preceding paragraph shall, after his death, be maintained, at least until the expiry of the economic rights, and shall be exercisable by the persons or institutions authorized by the legislation of the country where protection is claimed. However, those countries whose legislation, at the moment of their ratification of or accession to this Act, does not provide for the protection after the death of the author of all the rights set out in the preceding paragraph may provide that some of these rights may, after his death, cease to be maintained.

(3) The means of redress for safeguarding the rights granted by this Article shall be governed by the legislation of the country where protection is claimed” (OMPI, 1979).

Diante destas colocações, depreende-se que o Direito Autoral, mesmo enquanto reconhecimento do sustento legítimo de autores, só encontra justificativa plena na presença de limites ao seu escopo e exercício, balizado pela necessidade de resguardar direitos fundamentais (como o acesso à informação, à educação, à cultura e mesmo à justiça), bem como o atendimento de questões de interesse público, como o acesso à educação, ao conhecimento e à cultura. Estes limites configuram a base do sistema de limitações e exceções (L&E) ao Direito Autoral.

Ascensão (2010, p. 39) comenta que há uma diferença entre “limitações” e “exceções”: o segundo termo parte do princípio de que o direito à propriedade seria absoluto, tendo suas restrições caráter sempre excepcional. Já a noção de limitações, preferida pelo autor, deriva da ideia de que restrições são parte inerente a qualquer sistema de direitos, pela própria natureza da interação entre estes. Assim, Ascensão defende as limitações como parte constitutiva do Direito Autoral, em pé de igualdade com as regras que atribuem poderes aos titulares. Tal entendimento vai ao encontro do reconhecimento deste sistema como um mecanismo interno ao Direito de Autor, conforme já afirmavam A. Souza (2005, p. 242) e Carboni (2006, p. 97).

Mais do que isso, este trabalho compreende que a distinção entre ambos os termos reflete diferentes posições sobre o status da propriedade intelectual perante o patrimônio cultural e científico: “exceções” parece enfatizar a ideia de que o Direito Autoral emerge de uma condição natural e, portanto, autônoma em relação a outros direitos fundamentais; ele é a regra, aceitando apenas concessões pontuais. Já “limitações” evoca uma ideia de que os direitos do autor são parte integrante de um sistema pautado no incentivo à cultura e ao desenvolvimento científico e tecnológico – e, portanto, não existem de forma isolada, mas *em função* desse sistema. Em outras palavras, por esta acepção, o interesse público (e, portanto, o domínio público) é a regra e o Direito Autoral, a exceção – colocada diante de certas garantias e interesses individuais que servirão, no longo prazo, ao interesse público (Souza, A. 2005; Lewicki, 2007; Ascensão, 2010; Branco, 2011).

Este trabalho compreende que a segunda opção condiz melhor com as características legais do Direito de Autor e nos permite enxergar as limitações e exceções de forma ampla, dividindo-as em dois tipos principais: as intrínsecas, existentes na própria legislação autoral, e as extrínsecas, oriundas de exigências externas à estrutura do Direito Autoral que são ditadas não apenas por outros direitos fundamentais, mas por deveres do Estado e outras imposições

legais e sociais aos direitos de propriedade – formando um sistema onde o interesse do autor opera de forma integrada ao interesse público (Ascensão, 2010, p. 41; Carboni, 2006, p. 97).⁸¹

Mais especificamente, esta visão nos dá amplitude de enxergar as próprias limitações intrínsecas em diversos aspectos da regulação autoralista. A primeira delas, segundo Guilherme Carboni (2006, p. 97), diz respeito ao fato de o Direito Autoral ser restrito não só quanto ao seu objeto – como visto nos requisitos de originalidade e na distinção entre ideia e expressão, que discutimos em um momento anterior –, mas também quanto à sua duração: os direitos patrimoniais do titular de direitos não são perpétuos como um direito de propriedade convencional, mas perduram por um prazo definido em lei. Findo este prazo, a obra entra em domínio público, ocasião na qual pode ser utilizada sem a necessidade de autorização ou licença, desde que os direitos morais sejam respeitados. Na lei autoral brasileira, o prazo padrão é de 70 anos contados a partir de 1º de janeiro do ano subsequente à morte do autor.⁸² A terceira forma de limitação intrínseca são os usos livres, que constituem as possibilidades em que não é exigida a prévia autorização do titular para o uso de uma determinada obra protegida (Ascensão, 2010, p. 39). No Brasil, estes dispositivos estão previstos nos artigos 46 a 48 da lei autoral brasileira, em seu Capítulo IV – o qual trata especificamente das limitações ao Direito Autoral (Brasil, 1998b).⁸³

⁸¹ Um exemplo de limitação extrínseca é apresentado por Senftleben (2016, p. 19-20): o uso do Direito da Concorrência como mecanismo para limitar o poder de mercado dos titulares de direitos autorais sobre repositórios de conteúdo (como bancos de dados), em casos nos quais a função do Direito Autoral de incentivar a estimular a inovação e a cultura é melhor servida ao facilitar o acesso às obras para fins transformativos. É o exemplo de obrigar um titular a licenciar seu conteúdo quando há uma recusa injustificada em fazê-lo (Senftleben, 2016, p. 19-20).

⁸² “Art. 41. Os direitos patrimoniais do autor perduram por setenta anos contados de 1º de janeiro do ano subsequente ao de seu falecimento, obedecida a ordem sucessória da lei civil.

Parágrafo único. Aplica-se às obras póstumas o prazo de proteção a que alude o caput deste artigo” (Brasil, 1998b).

⁸³ “Art. 46. Não constitui ofensa aos direitos autorais:

I - a reprodução:

a) na imprensa diária ou periódica, de notícia ou de artigo informativo, publicado em diários ou periódicos, com a menção do nome do autor, se assinados, e da publicação de onde foram transcritos;

b) em diários ou periódicos, de discursos pronunciados em reuniões públicas de qualquer natureza;

c) de retratos, ou de outra forma de representação da imagem, feitos sob encomenda, quando realizada pelo proprietário do objeto encomendado, não havendo a oposição da pessoa neles representada ou de seus herdeiros;

d) de obras literárias, artísticas ou científicas, para uso exclusivo de deficientes visuais, sempre que a reprodução, sem fins comerciais, seja feita mediante o sistema Braille ou outro procedimento em qualquer suporte para esses destinatários;

II - a reprodução, em um só exemplar de pequenos trechos, para uso privado do copista, desde que feita por este, sem intuito de lucro;

III - a citação em livros, jornais, revistas ou qualquer outro meio de comunicação, de passagens de qualquer obra, para fins de estudo, crítica ou polêmica, na medida justificada para o fim a atingir, indicando-se o nome do autor e a origem da obra;

IV - o apanhado de lições em estabelecimentos de ensino por aqueles a quem elas se dirigem, vedada sua publicação, integral ou parcial, sem autorização prévia e expressa de quem as ministrou;

Internacionalmente, contudo, o debate em torno das limitações e exceções tem se debruçado com mais ênfase sobre a criação e implementação dos usos livres, razão pela qual estes dispositivos figuram no centro da discussão deste capítulo. Isto em si é reflexo de questões mais amplas acerca do papel dessas regulações e o que elas informam a respeito do sistema internacional de Direito Autoral – algo que merece atenção quando se pensa em construir um novo uso livre e que, portanto, procuramos trazer ao final deste capítulo.

2.3. O regime internacional de limitações e exceções

O principal dispositivo que trata das limitações e exceções em âmbito internacional é uma regra geral que serve como guia para a implementação de outras possibilidades de limitações: a chamada “regra dos três passos”, originalmente prevista no artigo 9, parágrafo 2 da Convenção de Berna, e presente em tratados posteriores.⁸⁴ Essencialmente, trata-se de uma diretriz-geral que permite a países-membros da Convenção estipular casos em que um determinado uso não viola os direitos de reprodução do titular, desde que em (1) certos casos especiais, (2) que não conflitem com a exploração normal da obra e (3) que não causem prejuízo injustificado aos interesses legítimos do autor.⁸⁵ Mais tarde, esta regra viria a ser expandida para

V - a utilização de obras literárias, artísticas ou científicas, fonogramas e transmissão de rádio e televisão em estabelecimentos comerciais, exclusivamente para demonstração à clientela, desde que esses estabelecimentos comercializem os suportes ou equipamentos que permitam a sua utilização;

VI - a representação teatral e a execução musical, quando realizadas no recesso familiar ou, para fins exclusivamente didáticos, nos estabelecimentos de ensino, não havendo em qualquer caso intuito de lucro;

VII - a utilização de obras literárias, artísticas ou científicas para produzir prova judiciária ou administrativa;

VIII - a reprodução, em quaisquer obras, de pequenos trechos de obras preexistentes, de qualquer natureza, ou de obra integral, quando de artes plásticas, sempre que a reprodução em si não seja o objetivo principal da obra nova e que não prejudique a exploração normal da obra reproduzida nem cause um prejuízo injustificado aos legítimos interesses dos autores.

Art. 47. São livres as paráfrases e paródias que não forem verdadeiras reproduções da obra originária nem lhe implicarem descrédito.

Art. 48. As obras situadas permanentemente em logradouros públicos podem ser representadas livremente, por meio de pinturas, desenhos, fotografias e procedimentos audiovisuais” (Brasil, 1998b).

⁸⁴ Tais como o Acordo TRIPS e o Tratado da OMPI sobre Direito de Autor (WCT).

⁸⁵ “Article 9 [Right of Reproduction: 1. Generally; 2. Possible exceptions; 3. Sound and visual recordings]

(...)

(2) It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author” (OMPI, 1979, p. 8).

incluir quaisquer direitos exclusivos, como pode ser visto no Tratado da OMPI sobre Direito de Autor⁸⁶ e no Acordo TRIPS.⁸⁷

Neste caso, segundo o Corpo de Apelações da OMC, a disposição de “certos casos especiais” implica em dizer que limitações e exceções devem ser precisamente delimitadas e distinguíveis entre si, sem que haja espaço para interpretações extensivas.⁸⁸ Por outro lado, isto não significa que um país precise se restringir a uma lista fechada de casos – apenas que cada novo uso live deve ser devidamente especificado.

Já “exploração normal” se refere especificamente a usos que tenham, ou possam vir a ter, considerável importância econômica ou prática para o titular – e isto não se estende a todos os usos comerciais possíveis de uma obra protegida, apenas aqueles que seriam práticas comuns para o titular (Senftleben, 2004, p. 181-182). Intui-se alguns exemplos, como a publicação de uma foto em uma revista, venda de produtos com a imagem de uma pintura, produção e distribuição de cópias de uma mídia (como discos e DVDs), entre outros. Lewicki (2007, p. 93) faz uma comparação com o cálculo dos lucros cessantes em matéria de reparação de danos no Direito Civil brasileiro: computa-se apenas o valor do lucro que seria razoável esperar caso o ilícito não houvesse ocorrido.

O terceiro passo, contudo, é desenhado de forma mais vaga, o que requer uma análise mais cautelosa. Mais uma vez, recorreremos à conceituação fornecida pelo Corpo de Apelações da OMC (Organização Mundial do Comércio, 2024, p. 6),⁸⁹ cujo entendimento, em suma, é o

⁸⁶ “Article 10

Limitations and Exceptions

(1) Contracting Parties may, in their national legislation, provide for limitations of or exceptions to the rights granted to authors of literary and artistic works under this Treaty in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author.

(2) Contracting Parties shall, when applying the Berne Convention, confine any limitations of or exceptions to rights provided for therein to certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author” (OMPI, 1996).

⁸⁷ “Article 13

Limitations and Exceptions

Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder” (ORGANIZAÇÃO MUNDIAL DO COMÉRCIO, 1994).

⁸⁸ “In our view, the first condition of Article 13 requires that a limitation or exception in national legislation should be clearly defined and should be narrow in its scope and reach. On the other hand, a limitation or exception may be compatible with the first condition even if it pursues a special purpose whose underlying legitimacy in a normative sense cannot be discerned. The wording of Article 13’s first condition does not imply passing a judgment on the legitimacy of the exceptions in dispute. However, public policy purposes stated by law-makers when enacting a limitation or exception may be useful from a factual perspective for making inferences about the scope of a limitation or exception or the clarity of its definition” (ORGANIZAÇÃO MUNDIAL DO COMÉRCIO, 2024, p. 3).

⁸⁹ “The Panel then proceeded to examine each of these terms in turn and began with their ordinary meaning: “The ordinary meaning of the term ‘interests’ may encompass a legal right or title to a property or to use or benefit of a

de que o terceiro passo prevê que uma limitação deve se dedicar a um uso que não interfere com o uso e usufruto geral (econômico ou não) de um bem jurídico garantido ao titular de direitos – ou, se o faz, tal uso se dá dentro de limites considerados justificáveis pela lei, como, por exemplo, o atendimento a questões de preponderante interesse público, como o acesso à educação. Isto pode incluir, por exemplo, direitos morais.

Como forma de clarificar os preceitos da regra dos três passos e sua aplicação de maneira adequada aos fins para os quais se destina, um grupo de acadêmicos de diversas instituições se uniu para a realização de uma Declaração, intitulada “Uma Interpretação Equilibrada do ‘Teste dos Três Passos’ do Direito de Autor”, estabelecendo que os três passos devem ser considerados de forma conjunta e que as limitações e exceções de Direito Autoral não podem ser interpretadas de forma restritiva, mas sim conforme seus propósitos. Defende-se a adoção de cláusulas abertas (desde que seu escopo seja razoavelmente previsível) e a possibilidade de os tribunais estabelecerem novas limitações. O documento também defende que as limitações não conflitam com a exploração normal da obra quando fundadas em “interesses concorrenciais importantes” ou servirem como oposição à aplicação de restrições abusivas à concorrência, especialmente em mercados secundários – o que é essencial para a mineração de dados, especialmente quando aplicada para fins comerciais. Finalmente, o documento encerra afirmando que, embora a regra dos três passos deva considerar os interesses dos titulares originários e subsequentes dos direitos autorais, ela também deve ser interpretada de forma a respeitar os interesses legítimos de terceiros, como os relacionados aos direitos humanos e ao progresso científico, cultural, social e econômico (Geiger *et al*, 2010).

O tópico das limitações e exceções nas organizações internacionais se estende para muito além do simples estabelecimento da Regra dos Três Passos, contudo. O Comitê Permanente sobre Direito Autoral e Direitos Conexos (SCCR), promovido pela Organização Mundial da Propriedade Intelectual (OMPI), mantém as L&E como parte permanente de sua agenda (OMPI, 2024a). Tais debates já encontravam proponentes fortes no Sul Global em 2004, quando

property (including intellectual property). It may also refer to a concern about a potential detriment or advantage, and more generally to something that is of some importance to a natural or legal person. Accordingly, the notion of 'interests' is not necessarily limited to actual or potential economic advantage or detriment. The term 'legitimate' has the meanings of '(a) conformable to, sanctioned or authorized by, law or principle; lawful; justifiable; proper; (b) normal, regular, conformable to a recognized standard type.' Thus, the term relates to lawfulness from a legal positivist perspective, but it has also the connotation of legitimacy from a more normative perspective, in the context of calling for the protection of interests that are justifiable in the light of the objectives that underlie the protection of exclusive rights. We note that the ordinary meaning of 'prejudice' connotes damage, harm or injury. 'Not unreasonable' connotes a slightly stricter threshold than 'reasonable'. The latter term means 'proportionate', 'within the limits of reason, not greatly less or more than might be thought likely or appropriate', or 'of a fair, average or considerable amount or size.'" (ORGANIZAÇÃO MUNDIAL DO COMÉRCIO, 2024, p. 6).

a Agenda do Desenvolvimento foi proposta na OMPI pelas delegações brasileira e argentina com o propósito de contextualizar o sistema de propriedade intelectual dentro de uma agenda de desenvolvimento socioeconômico mais equitativo entre os Países-Membros da Organização, levando em conta as particularidades de cada país e as flexibilidades conferidas pelo Acordo TRIPS (documento WO/GA/31/11) (OMPI, 2004). Em 2007, a Assembleia Geral da OMPI aprovou uma série de 45 recomendações para efetivação da Agenda, as quais incluem ações voltadas para “estabelecimento de normas, flexibilidades, políticas públicas e domínio público” (OMPI, 2024b).⁹⁰

Mais especificamente, a entrada do tópico como pauta do Comitê foi proposta em 2005, a partir de uma proposta feita pelo Grupo de Países da América Latina e Caribe (GRU-LAC), representado pela delegação do Chile (documento SCCR/13/5). O texto ressalta a importância das limitações e exceções como instrumento de balanceamento dos interesses de autores, titulares e usuários, bem como a necessidade de um estudo mais aprofundado sobre quais modelos de limitações existem entre os Países-Membros da OMPI e quais usos livres são necessários para a preservação do interesse público e o estímulo à criação e à inovação. A intenção seria estabelecer, ao fim, um instrumento multilateral para o estabelecimento de um regime de L&E mínimo para os Países-Membros da Organização (OMPI, 2005).⁹¹

Posteriormente, em 2009, esta pauta se intensificou no Comitê a partir de uma proposta preparada pela World Blind Union e apresentada pelas delegações do Brasil, Chile e Paraguai (documento SCCR/18/5) no sentido de estabelecer um regime internacional mínimo que permitisse a produção de versões de obras existentes para deficientes visuais (OMPI, 2009). Dois anos depois, um documento posterior do African Group (documento SCCR/22/12) incorporou um dispositivo similar, propondo também outras limitações, incluindo a possibilidade de

⁹⁰ Para mais informações, ver: <https://www.wipo.int/ip-development/en/agenda/recommendations.html>. Acesso em 15 fev. 2024

⁹¹ “At the twelfth session of the Standing Committee on Copyright of the World Intellectual Property Organization, Chile identified the relevance of prioritizing and setting aside working time for this Committee to strengthen international understanding of the need to have adequate limitations, learning from existing models and moving towards agreement on exceptions and limitations for public interest purposes which, like minimum standards, were to be envisaged in all legislations for the benefit of the international community.

On the basis of that statement, Chile wished to suggest three areas of work to be undertaken in the Standing Committee on Copyright and Related Rights:

1. Identification, from the national intellectual property systems of Member States, of national models and practices concerning exceptions and limitations.
2. Analysis of the exceptions and limitations needed to promote creation and innovation and the dissemination of developments stemming therefrom.
3. Establishment of agreement on exceptions and limitations for purposes of public interest that must be envisaged as a minimum in all national legislations for the benefit of the community; especially to give access to the most vulnerable or socially prioritized sectors.” (OMPI, 2005).

aquisição e compartilhamento de obras por arquivos e museus e, especialmente, à reprodução de obras para fins de ensino e pesquisa sem intuito de lucro, bem como L&E para o uso de programas de computador para fins de interoperabilidade e *backup* (OMPI, 2011).

A pauta de limitações para deficientes visuais encontrou mais avanços, culminando no *Tratado de Marrakesh para Facilitar o Acesso a Obras Publicadas às Pessoas Cegas, com Deficiência Visual ou com Outras Dificuldades para Ter Acesso ao Texto Impresso*, o primeiro tratado internacional específico para tratar de limitações e exceções. O preâmbulo do texto destaca o papel duplo do Direito Autoral enquanto recompensa pela criação artística e literária e, ao mesmo tempo, um instrumento para fomentar o acesso universal à cultura. Nesse sentido, as L&E são reafirmadas enquanto ferramentas para balancear a proteção efetiva dos direitos dos titulares com outras questões de interesse público, em especial o direito à educação e à pesquisa, bem como o acesso ao conhecimento (OMPI, 2013).⁹²

O Brasil confirmou sua posição como um dos proponentes originais do Tratado de Marrakesh ao ratificá-lo em 2015, incorporando-o ao ordenamento jurídico pátrio por meio de uma Emenda à Constituição, seguindo os trâmites previstos pelo atual regime constitucional. Desta forma, uma nova limitação específica de Direito Autoral foi adicionada ao rol de usos

⁹² “The Contracting Parties,

(...)

Mindful of the challenges that are prejudicial to the complete development of persons with visual impairments or with other print disabilities, which limit their freedom of expression, including the freedom to seek, receive and impart information and ideas of all kinds on an equal basis with others, including through all forms of communication of their choice, their enjoyment of the right to education, and the opportunity to conduct research,

Emphasizing the importance of copyright protection as an incentive and reward for literary and artistic creations and of enhancing opportunities for everyone, including persons with visual impairments or with other print disabilities, to participate in the cultural life of the community, to enjoy the arts and to share scientific progress and its benefits,

(...)

Recognizing both the importance of rightholders’ role in making their works accessible to persons with visual impairments or with other print disabilities and the importance of appropriate limitations and exceptions to make works accessible to these persons, particularly when the market is unable to provide such access,

Recognizing the need to maintain a balance between the effective protection of the rights of authors and the larger public interest, particularly education, research and access to information, and that such a balance must facilitate effective and timely access to works for the benefit of persons with visual impairments or with other print disabilities,

Reaffirming the obligations of Contracting Parties under the existing international treaties on the protection of copyright and the importance and flexibility of the three-step test for limitations and exceptions established in Article 9(2) of the Berne Convention for the Protection of Literary and Artistic Works and other international instruments,

(...)

Recognizing the importance of the international copyright system and desiring to harmonize limitations and exceptions with a view to facilitating access to and use of works by persons with visual impairments or with other print disabilities (...)” (OMPI, 2013).

livres do artigo 46,⁹³ tornando sua observação pelos contratos de direitos autorais obrigatória – sob pena de nulidade (Souza, A., 2016, p. 84-85; Souza, A e Fairbanks, 2016, p. 336-344).

No âmbito de limitações específicas para ensino e pesquisa, tais propostas não chegaram ao ponto de um dispositivo internacional, mas seguiram na pauta das reuniões da SCCR desde então. Inclusive, outras propostas nesse sentido foram apresentadas por países da América Latina: o documento SCCR/24/6, específico para educação, proposto pelas delegações do Equador, Peru e Uruguai (OMPI, 2012a), e uma proposta da delegação do Brasil (SCCR/24/7) com rascunhos de artigos e grupos temáticos referentes a limitações e exceções para instituições de ensino e pesquisa (OMPI, 2012b). Em 2018, foi estabelecido um plano de ações (SCCR/36/7) no sentido de explorar a situação de arquivos, museus e bibliotecas e os atuais modelos de L&E destinados a essas instituições (OMPI, 2018).

O Comitê também foi responsável pela Conferência Internacional sobre Limitações de Direito Autoral para Bibliotecas, Arquivos, Museus e Instituições Educacionais e de Pesquisa. Dentre os temas, foram incluídas questões relacionadas à preservação, acesso, cópias, uso transfronteiriço e outros usos, incluindo comerciais, de diversos acervos, visando a criação, no futuro, de uma nova disposição internacional para limitações visando estas instituições (OMPI, 2019a). Nesse ano, também foram trazidos estudos resultantes de um plano de ações, incluindo uma tipologia de limitações para arquivos (SCCR/39/5) (Crews, 2019) e um relatório (SCCR/39/6) sobre práticas e desafios no ensino à distância e na pesquisa (Torres e Xalabarder, 2019).

Posteriormente, os Estados Unidos apresentaram uma proposta de atualização do documento “OBJECTIVES AND PRINCIPLES FOR EXCEPTIONS AND LIMITATIONS FOR LIBRARIES AND ARCHIVES”, no qual encoraja a criação de limitações e exceções que permitam a condução das atividades de repositórios públicos enquanto instrumentos de preservação do conhecimento (Estados Unidos, 2023a).⁹⁴ Também em 2023, atendendo-se a uma proposta do African Group, foi implementado um programa de trabalho sobre limitações e

⁹³ “Art. 46. Não constitui ofensa aos direitos autorais:

I - a reprodução:

(...)

d) de obras literárias, artísticas ou científicas, para uso exclusivo de deficientes visuais, sempre que a reprodução, sem fins comerciais, seja feita mediante o sistema Braille ou outro procedimento em qualquer suporte para esses destinatários;” (Brasil, 1998b).

⁹⁴ “Exceptions and limitations for libraries, archives, and museums help individuals seeking, obtaining, and disseminating information, including cultural, artistic, and scientific knowledge, so that such individuals may participate meaningfully in public life. Exceptions and limitations advance knowledge, creativity, and innovation by preserving and providing access to the world’s cultural, artistic, and scientific heritage.” (Estados Unidos, 2023a).

exceções (OMPI, 2023c; 2023d), incluindo a facilitação da pesquisa por meios digitais⁹⁵ e destacando a convergência deste programa com o cumprimento do papel da OMPI no contexto da Agenda do Desenvolvimento das Nações Unidas.⁹⁶ Em 2024, o esboço do plano de implementação deste grupo de trabalho (SCCR/45/10 PROV.), a pedido do GRULAC, mencionou expressamente a mineração de dados como ligada à necessidade de se pensar em L&E para pesquisa, ensino e aprendizado no ambiente digital (OMPI, 2024c).⁹⁷

De volta a 2023, em uma reunião de outro órgão da OMPI – o Comitê sobre Desenvolvimento e Propriedade Intelectual (CDIP) –, o African Group também apresentou uma proposta de projeto voltada para a regulação da mineração de textos e dados para fins de pesquisa dentro do continente africano. A proposta (CDIP/30/9 REV.) foi revisada em abril do mesmo ano, concentrando-se, em grande medida, sobre o objetivo de identificar regulações e práticas de licenciamento para mineração de textos e dados em países africanos, bem como auxiliar na

⁹⁵ “This proposal is presented in furtherance of the implementation of the Work Program especially as it relates to the following three priority issues on exceptions and limitations highlighted in Paragraph/Point 2: (...) b. to promote the adaptation of exceptions and limitations to the online environment, such as by permitting teaching, learning and research through digital and online tools (...)” (OMPI, 2023c, p. 2).

⁹⁶ “The proposed work program has links to DA recommendations 1, 7, 9, 14, 32, 40 and 44:

1. WIPO technical assistance shall be, *inter alia*, development-oriented, demand driven and transparent, taking into account the priorities and the special needs of developing countries, especially LDCs, as well as the different levels of development of Member States and activities should include time frames for completion. In this regard, design, delivery mechanisms and evaluation processes of technical assistance programs should be country specific.

7. Promote measures that will help countries deal with intellectual property-related anticompetitive practices, by providing technical cooperation to developing countries, especially LDCs, at their request, in order to better understand the interface between IPRs and competition policies.

9. Request WIPO to create, in coordination with Member States, a database to match specific intellectual property-related development needs with available resources, thereby expanding the scope of its technical assistance programs, aimed at bridging the digital divide.

14. Within the framework of the agreement between WIPO and the WTO, WIPO shall make available advice to developing countries and LDCs, on the implementation and operation of the rights and obligations and the understanding and use of flexibilities contained in the TRIPS Agreement.

32. To have within WIPO opportunity for exchange of national and regional experiences and information on the links between IPRs and competition policies.

40. To request WIPO to intensify its cooperation on IP related issues with United Nations agencies, according to Member States’ orientation, in particular UNCTAD, UNEP, WHO, UNIDO, UNESCO and other relevant international organizations, especially the WTO in order to strengthen the coordination for maximum efficiency in undertaking development programs.

44. In accordance with WIPO’s member-driven nature as a United Nations Specialized Agency, formal and informal meetings or consultations relating to norm-setting activities in WIPO, organized by the Secretariat, upon request of the Member States, should be held primarily in Geneva, in a manner open and transparent to all Members. Where such meetings are to take place outside of Geneva, Member States shall be informed through official channels, well in advance, and consulted on the draft agenda and program.” (OMPI, 2023c, p.4).

⁹⁷ “[III. Adaptation to the online environment

Facilitate further work to promote the adaptation of limitations and exceptions to the online environment, such as by permitting teaching, learning and research through digital and online tools, due to its relationship with data mining.]” (OMPI, 2024c).

capacitação de instituições de pesquisa para empregar tais tecnologias de forma mais efetiva e legalmente segura (OMPI, 2023b).⁹⁸

Meses depois, em novembro de 2023, a sessão de número 44 da SCCR recebeu um estudo da professora Raquel Xalabarder sobre os desafios enfrentados por instituições de pesquisa no âmbito do Direito Autoral. Mais especificamente – e de forma similar à observada em outro trabalho (OMPI, 2019a) –, o estudo aponta para a defasagem de muitos regimes de limitações e exceções nacionais que permitem apenas a reprodução de material para fins de pesquisa, quando o ambiente digital não só impõe dificuldades ao acesso lícito, mas também implica a comunicação ao público em meio aos atos necessários para divulgar os resultados de pesquisa.⁹⁹ Para além de destacar a necessidade de se pensar na digitalização, adaptação e compartilhamento como atos intrínsecos à pesquisa, o estudo também ressalta o crescimento na importância da mineração de textos e dados e sua inclusão em regimes de licenciamento por editoras, assim como o crescimento de modelos de ciência aberta por instituições de pesquisa (Xalabarder, 2023).

A mineração de textos e dados também tem aparecido de forma notadamente expressa em sessões especiais da OMPI, intituladas de “Conversações da OMPI sobre Propriedade Intelectual e Inteligência Artificial”.¹⁰⁰ Sua primeira sessão se deu em 2019 – ainda com um viés mais otimista em relação às aplicações da inteligência artificial, concentrando-se em desenvolver ferramentas de IA e resolver possíveis *gaps* regulatórios entre Países-Membros (OMPI,

⁹⁸ “This project focuses on TDM from the perspective of universities and other research-oriented institutions in Africa. Universities and research-oriented institutions have a key role to play as stakeholders in the TDM landscape. In depth mapping of copyright treaty implementation, the copyright legal framework including specific provisions for research, as well as licensing arrangements and practices, is needed for African universities and other research-oriented institutions to effectively access TDM in a useful way.

The proposed pilot project will promote awareness and build capacities to use TDM by universities and other research-oriented institutions in Africa and thus facilitate the use of AI tools, create and disseminate knowledge on TDM through documenting best practices by universities or research institutions in the region, and build capacities of staff from universities and research institutions in Africa.” (OMPI, 2023b, p. 1-2).

⁹⁹ “Historically, research activities were carried out in academic institutions and universities, mainly involving the creation of hard copies of specific library collections. These actions focused primarily on published works and, to some extent, on the reproduction of artistic works, such as films or songs, within the context of research. To a large extent, these uses were limited to the right of reproduction, especially in analog formats such as paper copies or photocopies. As a result, many national legislations only mentioned reproduction and communication to the public in analog contexts and in relation to research institutions.

In contrast, when research is conducted online and through digital means, it is not limited only to reproduction, but also involves communication to the public and making it available to the public.

(...)

Researchers share copies with each other, either directly from one researcher to another or through shared access drives and internal networks. These actions may extend through the Internet and social networks. However, the categorization of these activities in the realm of distribution and communication to the public may change as the size of the team and the audience reached increases.” (OMPI, 2023a, p. 8).

¹⁰⁰ A página principal com links para todas as sessões se encontra disponível em: https://www.wipo.int/about-ip/en/artificial_intelligence/conversation.html. Acesso em 20 set. 2024.

2019b). Já na Segunda Sessão, já se observa uma preocupação maior com a possibilidade de prejuízo aos titulares pela IA Generativa, ao mesmo tempo em que há menção explícita à necessidade de se pensar em hipóteses de usos livres de mineração de textos e dados (incluindo para pesquisa); de se considerar como a proibição do uso de obras em *machine learning* afeta as limitações e exceções para mineração de textos e dados já em vigor e, finalmente, como realizar o licenciamento de obras protegidas para fins de treinamento de IA (OMPI, 2020).¹⁰¹

Finalmente, a SCCR/44 também observou uma proposta para sessões sobre IA Generativa, apontando a questão de se observar o uso de obras protegidas para o treinamento de inteligências artificiais (OMPI, 2023d).¹⁰² Tal proposta culminou na realização de uma Sessão Informativa sobre IA e Direito Autoral durante a 45ª Sessão do Comitê em abril de 2024, da qual participaram representantes da classe artística, do setor jornalístico, empresas de tecnologia, academia e museus, entre outros. A variedade de palestrantes também se refletiu em relação aos posicionamentos gerais sobre a inteligência artificial generativa: uns manifestaram apoio ao desenvolvimento de tecnologias de IA, outros se posicionaram contra, aceitando apenas o uso de coleta e análise para fins de pesquisa e educação, e houve posicionamentos de maior rechaço ou desconfiança à IA e aos procedimentos eletrônicos de coleta e análise de dados de forma geral – e, conseqüentemente, às próprias limitações e exceções de direitos autorais já destinadas para tais fins (OMPI, 2024d).¹⁰³

¹⁰¹ “Issue 8: Infringement and Exceptions

24. “An AI application can generate creative works by learning from data with AI techniques such as machine learning. The data used for training the AI application may represent creative works that are subject to copyright (see also Issue 11). A number of issues arise in this regard, specifically, (...)

(iii) If the use of data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an explicit exception be made under copyright law or other relevant laws for the use of such data to train AI applications?

(iv) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an exception be made for at least certain acts for limited purposes, such as the use in non-commercial user-generated works or the use for research?

(v) If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, how would existing exceptions for text and data mining interact with such infringement?

(vi) Would any policy intervention be necessary to facilitate licensing if the unauthorized use of data subsisting in copyright works for machine learning is considered an infringement of copyright? Would the establishment of mandatory collective management societies facilitate this? Should remedies for infringement be limited to equitable remuneration?” (OMPI, 2020).

¹⁰² “Stakeholders of all creative industries around the world are concerned with the fundamental impact of generative AI. While acknowledging new opportunities to accelerate and simplify work processes, they point to considerable risks for creative diversity and democratic opinion-forming. Above all, a large part of the creative sector sees the foundations of its business profoundly affected, since works created by humans could be partially or even completely replaced by AI-generated content.

The most discussed issues regarding copyright law include questions in relation with: *the use of copyright protected content as training data for AI models and applications (...)*” (OMPI, 2023f. Destaque feito pelo autor).

¹⁰³ O vídeo da sessão, bem como a transcrição da mesma, se encontram disponíveis em: https://webcast.wipo.int/video/SCCR_45_2024-04-18_PM_122422. Acesso em 6 set. 2024.

Vê-se, portanto, que o espaço destinado à mineração de textos e dados no âmbito do Direito Autoral em canais formais a nível global ainda é recente. Não se vislumbra, portanto, qualquer possibilidade de um instrumento internacional específico para a mineração de textos e dados na pesquisa científica – ficando a regulamentação desta questão a cargo de entidades nacionais e regionais, como as que veremos no Capítulo seguinte.

Por outro lado, a entrada da IA Generativa na pauta de discussões da Organização Mundial da Propriedade Intelectual chama atenção para o fato de que o processo de tramitação da pauta das limitações e exceções para fins de pesquisa, ensino e conservação no ambiente digital – incluindo a mineração de textos e dados – agora passa a conviver com outra discussão que parece abordar as tecnologias cobertas por essas limitações de maneira mais receosa. Parece relevante, portanto, compreender o que isto representa para o regime de L&E – e, especialmente, como isto afeta a construção dos usos livres atualmente postos (ou propostos) para a mineração de textos e dados em atividades de pesquisa ao redor do mundo.

2.4. O papel de um uso livre para mineração de textos e dados no atual sistema de limitações e exceções

Ao longo deste Capítulo, procurou-se oferecer um panorama geral acerca das limitações e exceções ao Direito Autoral e seu papel enquanto expressão da necessidade de adequar o instituto à sua função enquanto meio de fomento à produção de conhecimento e seu usufruto pela sociedade – e, nesse sentido, o reconhecimento da mineração de textos e dados para fins de pesquisa como ferramenta importante no atual processo de desenvolvimento cultural, científico e tecnológico. Contudo, ainda é pertinente colocar alguns questionamentos finais, especialmente em face das problemáticas trazidas pela IA Generativa vistas no Capítulo 1 e na seção anterior. Mais especificamente, cumpre observar de que forma as L&E vêm sendo mobilizadas em relação ao sistema de Direito Autoral como um todo e o que isso representa para as regulações sobre mineração de textos e dados, especialmente quando estas são tensionadas com pressões para uma regulação mais restritiva sobre tecnologias que dependem dessas ferramentas.

Nesse sentido, uma parcela da literatura de Direito Autoral sugere que pensar um modelo para qualquer uso livre novo (como no caso da mineração de textos e dados) requer uma reflexão acerca do próprio fundamento das limitações e exceções como forma de “equilibrar” ou “balancear” o direito de autor com outros direitos culturais ou informacionais e como o ambiente político e institucional, especialmente em âmbito internacional, influencia a efetivação desse objetivo. Para esses teóricos, no entanto, este “balanceamento”, embora explícito na teoria, estaria severamente desajustado na prática, em prejuízo dos usuários – levando a

situações como a “caça às exceções” descrita por Ascensão (2002) na década de 1990, onde a expansão do escopo de proteção do Direito Autoral não teria sido acompanhada por disposições pró-público na mesma intensidade.

Para alguns, as origens deste descompasso se localizam dentro da própria base teórica do sistema de propriedade intelectual. Perlman (2000), por exemplo, aponta que os direitos de propriedade encontram um arcabouço teórico muito mais “concreto” quanto aos seus benefícios para a inovação, enquanto a importância do compartilhamento e do acesso permaneceria em uma posição mais vaga e, portanto, de desvantagem argumentativa.¹⁰⁴ Ao mesmo tempo, Lemley nota uma tendência de diversos acadêmicos em se deslocar para uma visão de propriedade intelectual como um fim em si mesmo ao invés de pautada em seus efeitos práticos – o que o autor chama de “PI baseada na fé” (Lemley, 2015). Consequentemente, Craig (2017) nota a prevalência de um sistema que, em última análise, ainda tem em sua base o direito de autor como natural, como a “norma”. Neste cenário, as limitações e exceções – e, por extensão, o interesse público – assumiriam um papel secundário, posterior, manifestando-se como uma invasão aos interesses legítimos do autor.¹⁰⁵

Nesse sentido, Okediji (2018) aponta que este problema está intrinsecamente relacionado com a manutenção de um regime de propriedade intelectual ainda profundamente moldado por uma base filosófica essencialmente liberal e de pretensão universalizante, que privilegia valores e questões predominantes no Norte Global e que se concentra sobre usuários e atos individuais, sendo inadequada para necessidades institucionais ou coletivas – questões de maior interesse no Sul Global.¹⁰⁶ Nesse sentido, para Vyas (2023), isto é agravado pelo fato de que o regime internacional de Direito Autoral teria sido concebido dentro de um sistema

¹⁰⁴ “At what point does the balance between protection and access result in a net reduction of innovative activity? If we have no solid ground upon which to answer this question, should we be surprised if the claims for expanding property rights continue to erode the claims of those who ask for the right to “reap where they have not sown?”. In the contest between property rights and access rights, property rights always have the home field advantage. The incentives created by property rights are clear and the rhetoric is powerful.” (Perlman, 2000).

¹⁰⁵ “When authorial right is a baseline assumption, copyright exceptions or limitations are inevitably viewed with suspicion, manifesting as prima facie unjust encroachments upon the natural entitlement of the worthy, rights-bearing author.” (Craig, 2017, p. 12). Craig, C. J. ‘Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks’. 33 Am. U. Int’L Rev. 2, 2017.

¹⁰⁶ “Despite recognition of the many challenges imposed by IP rules, with dogged persistence, scholars along with national and international policymakers continue to defer to the notion of ‘balance’; balance within the IP system—and between IP and human rights—as the best response to the human welfare costs of global IP rules (...). Limitations and exceptions to IP rights certainly can address specific challenges, but rarely are they sufficient to meet the development-related challenges—such as bulk access to educational works—facing many least-developed and developing countries (...). Existing limitations and exceptions available in international copyright law, and in many domestic copyright laws, do not extend to institutional, community or group needs.” (Okediji, 2018, p. 21-23).

internacional colonialista, construído sem a participação de grande parte dos países que hoje compõem o Sul Global.

Nestes termos, para Okediji (2018), as L&E – e, em especial, os usos livres – tenderiam a ser situadas como a única maneira disponível e necessária de se regular o interesse público de acesso à cultura, ao conhecimento e aos produtos destes, limitando a capacidade dos Estados de pensar modos alternativos de regular a propriedade intelectual e, principalmente, de ir além dela.¹⁰⁷ Na prática, isto implica em um uso das limitações e exceções que as considera não como componente-padrão dentro do sistema de direitos autorais, mas como mecanismos de contenção de excessos (Okediji, 2014)¹⁰⁸ – ou, talvez, de contenção de danos. Para Vyas, isto se traduziria em um uso míope das limitações e exceções, que termina por colocar as L&E não como ferramenta de socialização do Direito Autoral, mas como um “fiador” que legitima expansões nos direitos dos titulares com a falsa promessa de que o balanceamento de direitos seria contramedida suficiente.

Ao mesmo tempo, Vyas (2023, p. 106) nota que a ligação entre “limitações e exceções” à noção de “balanceamento de direitos” carrega consigo a implicação de que qualquer uso não-fundamentado em um direito fundamental específico se tornaria, automaticamente, uma violação de Direito Autoral.¹⁰⁹ Uma consequência disto em limitações para a mineração de textos e dados na ciência, por exemplo, seria a de que simplesmente fundamentá-la em um “direito à pesquisa” poderia levar à interpretação de que quaisquer práticas de coleta e análise de dados, mesmo que não prejudiquem os interesses dos titulares, serão tratados como ilícitos

¹⁰⁷ “There are at least three ways that a broadened vision of the IP/human rights interface might produce a richer discourse for IP reform: 1) economic, social and cultural rights offer room to argue for doctrines that allow for use of knowledge goods at a greater *scale* than IP law currently allows, thus impacting a greater number of users; 2) economic, social and cultural rights could facilitate rules that require recognition of “users rights” and requires States to *ensure justiciability* of these rights in domestic law; 3) economic, social and cultural rights could be used to require State accountability for its choices with respect to the *implementation* of international IP obligations. Indeed, the ideals expressed in economic, social and cultural rights may require States to consider more seriously new models of innovation, including new business models, that can ensure returns to creators and access to users.” (Ibid., p. 25).

¹⁰⁸ “At its core, the human rights [frame] does not encounter intellectual property as much as it complements it; it does not resist it as much as it attempts to contain its effects. The . . . [IP/human rights interface] frames human rights as a redemptive discipline important to the mission of recognizing and protecting the ability of people to participate in the public interest goals animating the intellectual property system. The point of the human rights [frame] is not to transform intellectual property rights . . . [i]n indeed, the human rights [frame] sees the intellectual property system as a constitutive part of the universal values it seeks to affirm in and among states” (Okediji, 2014, p. 361).

¹⁰⁹ “(F)air dealing is meant to (though not limited to) serve that purpose i.e., to save one from copyright liability. The problem arises when fair dealing becomes a sheer matter of faith, where every copyright claim is expected to be controverted by fair dealing, where instead of examining whether the defendant’s use is infringing, the assumption is that the use is infringing causing the focus to remain on “escaping” liability via exceptions to copyright infringement which consequently develop a saviour-like image in our legal thought and become a defense against copyright” (Vyas, 2023, p.106).

a menos que possam se justificar neste direito, até mesmo para o acesso¹¹⁰ – que não constitui um direito exclusivo, mas aparece como exigência nos textos de diversas regulações nesse sentido. Efetivamente, portanto, o Direito Autoral se torna a régua para o interesse público, ao invés de lhe prestar deferência.

Ainda, para o autor, há outro agravante: historicamente, a ideia de “balancear direitos” nunca teria sido formalmente aprofundada e delimitada, permanecendo aberta a diferentes interpretações e aplicações conforme os interesses de políticas públicas vigentes em uma dada época (*Ibid.*). Por um lado, isto chegou a gerar resultados favoráveis à mineração de textos e dados: a crescente presença dessas tecnologias no processo de inovação ao longo das duas últimas décadas tem se beneficiado de um movimento de abertura do Direito Autoral, visivelmente pautado não somente em uma justificativa de atendimento a direitos fundamentais, mas também de incentivo à inovação e à pesquisa – o que se reflete na inclusão dessas tecnologias às diversas propostas de estudos e discussões sobre limitações e exceções na esfera supranacional. Por outro lado, conforme visto no Capítulo anterior, a popularização de modelos de IA Generativa tem levado a crescentes pressões de titulares por uma regulação mais incisiva sobre o uso de obras protegidas em processos de coleta e análise automatizada de dados – o que também já se nota em diversas falas no âmbito da própria OMPI referentes à inteligência artificial (OMPI, 2024b). Neste cenário, portanto, receia-se que esta noção indefinida de “balanceamento” seja empregada para a imposição de regras que, no afã de limitar práticas antiéticas por empresas de tecnologia, terminem por impor restrições aos usos livres já existentes para a coleta e análise de dados – ou, no mínimo, por limitar o alcance de possíveis regulações do gênero em locais que ainda não as possuem, como no caso do Brasil.

Tudo isto coloca a construção de uma regulação para a mineração de textos e dados no regime de limitações e exceções em posição mais vulnerável, agravada pelo caráter vago da regra dos três passos (Senftleben, 2022). A exigência de “certos casos especiais”, por exemplo, seria nebulosa, considerando a quantidade de materiais normalmente necessários em uma análise de dados e o número restrito de beneficiários a serem alcançados por uma limitação do gênero. O mesmo pode ser dito da interferência na exploração normal e em interesses legítimos

¹¹⁰ “While defining public interest in terms of rights is over a century old, as clear from the German proposal discussed in the previous sections, in recent years, it has increased, especially in defining fair dealing.(...) It is especially relevant and more problematic for technologies that do not engage with copyrighted works in the traditional sense of “use” and “access”. The recent demand for the Right to Research (“R2R”) to legitimize Text and Data Mining (“TDM”) is an apt example in this regard.¹¹⁹ My thesis is that creating new rights or exceptions based on an R2R will generalize the Whole research and access issue in a way where “accessing” works necessarily results in its “use.” To legitimize such usage, it becomes necessary to create a new amendment or exception in copyright law, drawing R2R.” (Vyas, 2023, p. 110-111).

no que diz respeito ao uso de material para fins de pesquisa, criando um ambiente de incerteza em relação à gestão dos dados por parte de pesquisadores, jornalistas e profissionais de conservação.

Ao mesmo tempo, porém, o autor (*Ibid.*) destaca que a própria definição de reprodução no ambiente digital – que perpassa a mineração de textos e dados, como visto anteriormente – também se encontra incerta. Observa-se que diversos documentos envolvendo a OMPI parecem considerar que a reprodução e o armazenamento eletrônicos de obras necessitam apresentar as mesmas a usuários para que se configure um ato coberto pelos direitos exclusivos dos titulares. Ao mesmo tempo, diversas delegações se mostraram reticentes em ampliar o escopo dos direitos de reprodução no ambiente digital, o que levaria à conclusão apontada por outros autores (vide Sag, 2019; Carroll, 2020; Margoni e Kretschmer, 2022) de que a mineração de textos e dados tecnicamente constituiria um uso que sequer poderia ser considerado pacificamente dentro do escopo dos direitos exclusivos dos autores.¹¹¹ Logo, a criação de uma limitação poderia ter o efeito adverso de limitar desnecessariamente a inovação em prol de uma segurança jurídica que só atenderá a setores específicos.¹¹² Margoni e Kretschmer (2022) se posicionam de forma similar, apontando que, se a mineração de textos e dados por si só lida com o uso de dados – aqui entendidos como exclusivamente de natureza factual –, então um uso livre estritamente delimitado pode incorrer na invasão a princípios como a distinção ideia-expressão e a não-proteção a fatos. Em outras palavras, uma limitação ferindo outras.

Em face destas considerações, pode-se ter a impressão de que instituir novos usos livres representaria uma perda de tempo; quiçá, de que já que seria impossível reconciliá-los com o verdadeiro atendimento ao interesse público. O que se argumenta, contudo, é o oposto: a

¹¹¹ “In the 1982 WIPO/UNESCO Committee, there was thus support for a “functional” approach asking whether “the work can be perceived, reproduced or otherwise communicated to the public.” (...) the delegates at the Diplomatic Conference were not convinced that temporary copying should be qualified as a relevant act of reproduction. This outcome of the Conference was reached even though, for instance in cases of browsing, temporary copies enable an individual user to perceive and enjoy the work.

Against this background, there is even more reason to assume that the international right of reproduction laid down in Art. 9(1) BC does not apply when, from the outset, copies are not made for the purpose of allowing an individual to perceive, reproduce or otherwise communicate the work.¹¹⁹ In the case of TDM copies, works are only included in a harmonized dataset to allow the analysis with the help of an automated, computational process. The requirement of a reproduction allowing an individual to consult and enjoy the work is not fulfilled.” (Senftleben, 2022, p. 1497-1498).

¹¹² “In other words: Art. 9(1) BC cannot be understood to cover TDM until such time as an international consensus is reached on this very question. As the topic of TDM has not yet been discussed at an international conference clarifying the scope of the right of reproduction, current TDM regulations that presume copyright relevance and seek to prevent inroads into the right of reproduction, such as the restrictive approach taken in the EU, constitute Berne/TRIPS/WCT-plus regimes that have been adopted voluntarily in domestic legislation to offer more copyright protection than necessary under international law.” (*Ibid.*, p. 1502).

criação de novas hipóteses de usos permitidos constitui uma parte essencial do regime de limitações e exceções, mas tem sua efetividade fragilizada quando estas previsões legais são tratadas de forma secundária e em termos vagos, ao mesmo tempo em que são colocadas como única opção para o atendimento do interesse público no Direito Autoral, em detrimento de outros instrumentos similares na lei e de qualquer outra possibilidade regulatória nesse sentido. Consequentemente, tem-se um mecanismo de balizamento do Direito Autoral que, isoladamente, pouco pode fazer dentro de um sistema de proteção que se torna cada vez mais ilimitado.

Ao fim e ao cabo, o que se observa é que o sistema de limitações e exceções, enquanto mecanismo de adequação do Direito Autoral à sua função de fomentar a expansão do conhecimento enquanto bem coletivo, é tensionado contra um sistema de propriedade que ainda carrega consigo fundamentos teóricos centrados em um modelo de apropriação privatista, pautado em direitos naturais e absolutos – um cenário onde a busca pelo “equilíbrio” termina comprometida por um *desequilíbrio*, no qual os mecanismos de garantia da função social da propriedade em posição desvantajosa. No caso da mineração de textos e dados, isto se traduz na formação de um debate marcado por uma crescente tensão entre as necessidades dos profissionais de pesquisa e do ecossistema de inovação, de um lado, e renovadas pressões pelo fortalecimento dos direitos sobre obras protegidas, alimentadas por preocupações com a substituição do trabalho humano no campo da produção cultural e científica. Nesse cenário, a ausência de dispositivos robustos de limitações e exceções pode resultar em um arcabouço regulatório pouco compatível com uma distribuição justa de prerrogativas, com um efetivo equilíbrio de direitos e com as demandas do interesse público.

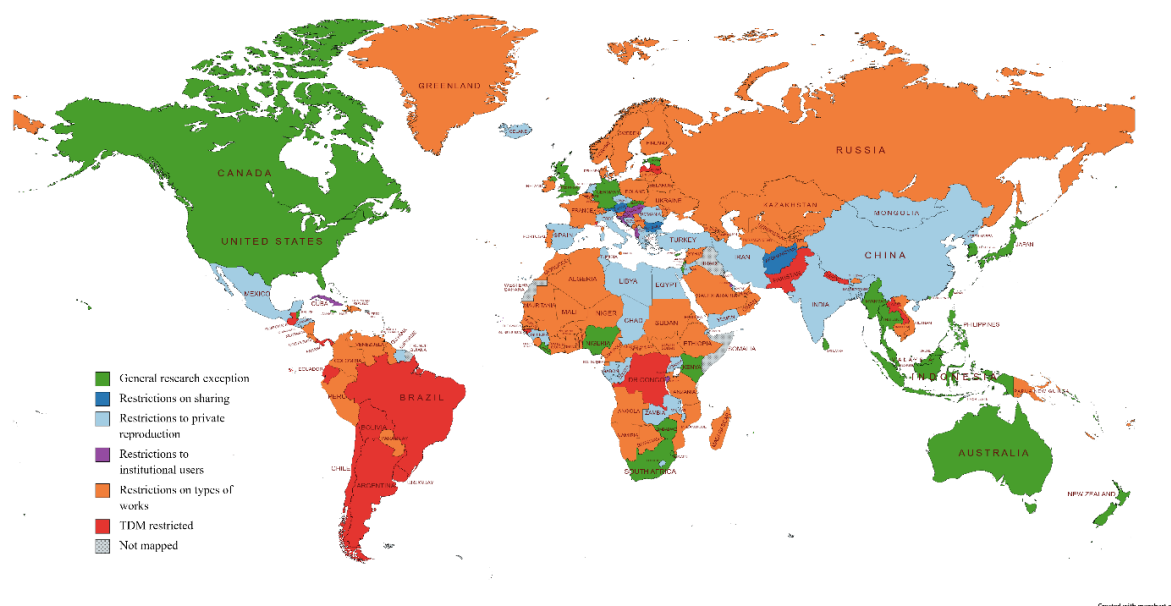
Resta observar de que maneiras as experiências nacionais e regionais ao redor do mundo têm lidado com esta tensão, especialmente em um momento no qual as estratégias de inovação passam a se chocar com um renovado receio quanto aos desdobramentos sociais e econômicos dos recentes avanços tecnológicos. E, acima de tudo, é importante compreender, através destas diferentes experiências regulatórias, como o Brasil – que também sofre influência desses choques – pode aprender com as mesmas, a fim de alcançar um resultado verdadeiramente satisfatório para o beneficiário definitivo do conhecimento: a coletividade.

CAPÍTULO 3. LIMITAÇÕES DE DIREITO AUTORAL PARA MINERAÇÃO DE TEXTOS E DADOS NAS EXPERIÊNCIAS NACIONAIS

Apresentamos, a seguir, o panorama geral das limitações e exceções para a mineração de textos e dados ao redor do mundo – não através de uma análise de direito comparado, mas sim enquanto política de inovação, com dois objetivos centrais. O primeiro é situar a regulação jurídica do tema enquanto política pública, ou, nos casos em que a regulação se dá no âmbito do Poder Judiciário, enquanto comumente inserida em um contexto mais amplo de inovação tecnológica. O segundo objetivo é ilustrar diferentes opções regulatórias e indicar pontos relevantes para consideração dentro da discussão regulatória sobre mineração de textos e dados no Brasil – alguns dos quais já haviam sido esboçados em um momento anterior (Alvarenga, 2019), mas que agora são aglutinados enquanto modelos regulatórios, visando uma quantidade maior de exemplos e uma visualização melhor das alternativas existentes.

Antes de passarmos para exemplos mais específicos, é relevante apresentar um panorama geral desses regimes ao redor do mundo – o que Flynn et al (2022) proporcionam a partir de seu levantamento acerca de regimes legais em todo mundo, classificando-os em um sistema de cores conforme o grau de permissividade para a limitação de textos e dados para fins de pesquisa científica – tendo em mente a existência de restrições a atos permitidos (reprodução e/ou compartilhamento), tipos de usuário, tipos de obra ou ao propósito da mineração (podendo ter uso comercial ou não). Assim, o trabalho divide os países em seis cores: vermelho (proibição total ou com restrições excessivas, como pequenos trechos ou inexistência de licença disponível); laranja (restrição a tipos de obras ou o quanto pode ser usado); roxo (apenas certos usuários, como instituições, podem reproduzir); azul (apenas reprodução, mas cópias privadas são possíveis); azul escuro (sem restrições, mas compartilhamento não é permitido) e verde (sem restrições, embora a proibição de uso comercial ainda possa existir). O resultado gráfico deste levantamento pode ser visto no mapa a seguir (Figura 2):

FIGURA 6: Mapa com classificação dos países conforme o grau de abertura de Direito Autoral para mineração de textos e dados na pesquisa



Fonte: Flynn et al, 2022.

Aqui, é possível notar uma certa correspondência a uma tendência regulatória a nível internacional, já apontada por Flynn e Palmedo (2017): ao compararem as regulações de Direito Autoral entre países, os autores constataram que havia um movimento geral de maior abertura ao longo do tempo – mas que, na média, havia um atraso significativo dos países do Sul Global em relação ao Norte. Assumindo-se a relação entre abertura e desenvolvimento científico e tecnológico apontada pelos autores no mesmo trabalho – conforme mencionado mais acima – então isto coincide com a percepção de que uma divisão internacional do trabalho que coloca países em desenvolvimento em desvantagem científica e tecnológica ainda persiste.

Entretanto, também é importante notar algumas especificidades indicadas pelo mapa, dado o caráter relativamente novo da mineração de textos e dados no ambiente regulatório internacional: uma parcela significativa dos BRICS – China, Índia e África do Sul – aparece com graus significativos de abertura à mineração de textos e dados para pesquisa, bem como outros países na África e Ásia (a exemplo de Nigéria, Quênia, Vietnã e Indonésia). É importante frisar que isto não corresponde a dizer que tais países já trataram diretamente sobre o tema, mas que certas possibilidades regulatórias preexistentes, especialmente no que se refere a permissões amplas para fins de pesquisa, podem ser consideradas para a coleta e análise de dados – conforme será discutido mais adiante.

O presente trabalho opta, ao invés de uma classificação baseada em abertura, por uma divisão baseada em modelos regulatórios, com base em semelhanças de abordagem. Contudo,

a classificação acima será utilizada como auxílio na observação do grau de permissividade que se permite estender a pesquisadores dedicados à análise de dados.

Também tomamos como suporte teórico o trabalho de Gervais (2019, p. 30-31), sobre tópicos centrais que costumam se fazer presentes nas limitações para mineração de textos e dados observadas pelo autor à ocasião. Resumidamente, o autor elenca os seguintes pontos:

1. Quais direitos exclusivos esta limitação cobre? Apenas reprodução, ou também adaptação/criação de obras derivadas, comunicação ao público, entre outros?
2. É possível renunciar a essa limitação pela via contratual?
3. O material utilizado deve ser proveniente de uma fonte lícita?
4. De forma relacionada à primeira pergunta: a disseminação desses dados seria possível? E de que forma?
5. A mineração de dados deve ser apenas para fins não comerciais?
6. Por fim, esta limitação permitiria “burlar” medidas tecnológicas de proteção ou apenas serviria como recurso legal contra medidas destinadas a restringir esta “burla”?

Tendo essas questões como guia, observaremos os diferentes modelos regulatórios sobre o assunto encontrados ao longo da pesquisa documental, assim como destacaremos pontos adicionais de interesse que venham a emergir da leitura dos textos legais. Ao mesmo tempo, contudo, é importante asseverar que tais questões foram, em grande parte, aplicadas a limitações e exceções contidas em textos legais, podendo necessitar de ajustes quando usadas em contextos jurídicos diferentes – tais como o modelo de uso justo, abordado a seguir.

3.1. Estados Unidos: o modelo de cláusula aberta e o uso justo (*fair use*)

O primeiro modelo de limitação de Direito Autoral para mineração de textos e dados a ser apresentado neste trabalho não é exatamente o primeiro a conferir caráter explicitamente lícito à atividade, mas as bases jurídicas adotadas para este fim precedem outras limitações e exceções sobre o tema existentes ao redor do mundo. Trata-se do modelo de cláusula aberta, que não apenas elenca alguns usos que não implicam em violação de Direito Autoral, como também confere às cortes o papel de estabelecer situações adicionais de não-violação para além do que consta no texto da lei. Um exemplo emblemático deste modelo é a doutrina do uso justo (*fair use*, no original em inglês), presente em diversos países, tais como Singapura (2021) (art.

190 e seguintes) e Coreia do Sul (2023) (art. 35.5)¹¹³ Entretanto, esta seção tomará como exemplo central os Estados Unidos, que já fizeram uso concreto da doutrina para permitir certos usos relacionados à mineração de textos e dados.

Neste contexto, a doutrina, que no Direito estadunidense surgiu nos tribunais, já é reconhecida em lei há décadas:¹¹⁴ a seção 107 do *Copyright Act* estabelece os critérios a serem considerados pelos tribunais: (i) o propósito e caráter do uso; (ii) a natureza da obra utilizada; (iii) a quantidade e relevância dos trechos usados em relação ao total da obra e, por fim, (iv) o impacto do uso sobre o mercado ou valor da obra.¹¹⁵

Em relação ao primeiro fator (o propósito e o caráter do uso), a questão essencial é se o uso da obra protegida por direitos autorais é transformador. Em outras palavras, conforme estabelecido no caso *Campbell v Acuff-Rose* (em relação a paródias), é preciso verificar se o uso novo substitui o uso padrão da obra original ou se ele carrega um novo propósito ou significado que o distingue suficientemente da obra.¹¹⁶ Quanto mais transformador for esse uso, menos peso será dado aos outros fatores.¹¹⁷ Outro ponto digno de nota estabelecido por *Campbell* é a necessidade do uso daquela obra para que se conclua pela transformação: qualquer uso

¹¹³ Tais países são apenas exemplos: Equador, Filipinas, Israel e Taiwan, dentre outros, também possuem cláusulas de uso justo (como pode ser visto, por exemplo, em Flynn et al, 2022).

¹¹⁴ “The judicial doctrine of fair use, one of the most important and well-established limitations on the exclusive right of copyright owners, would be given express statutory recognition for the first time in section 107. The claim that a defendant’s acts constituted a fair use rather than an infringement has been raised as a defense in innumerable copyright actions over the years, and there is ample case law recognizing the existence of the doctrine and applying it” (Estados Unidos, 1976b, p. 65).

¹¹⁵ “Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors”(Estados Unidos, 1976a).

¹¹⁶ “The central purpose of this investigation is to see, in Justice Story’s words, whether the new work merely ‘supersede[s] the objects’ of the original creation, *Folsom v. Marsh*, supra, at 348; accord, *Harper & Row*, supra, 471 U.S., at 562, 105 S.Ct., at 2231 (‘supplanting’ the original), or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is ‘transformative’” (Estados Unidos, 1994b, p. 579/1171).

¹¹⁷ “(...) the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use” (*Ibid.*).

de uma obra protegida por direitos autorais precisa ser feito na medida em que (e na medida em que) seja necessário para transmitir a mensagem desejada.¹¹⁸

Quanto ao segundo fator, (a natureza da obra), observa-se, por exemplo, se o material protegido utilizado constituía um trabalho mais expressivo (como uma ficção) ou se era uma descrição mais factual – sendo o uso no segundo caso mais favorável do que no primeiro.¹¹⁹ Ao mesmo tempo, observa-se se o trabalho havia sido tornado disponível ao público: em caso negativo, este elemento se torna relevante em desfavor de um uso justo¹²⁰ – ainda que, conforme dita a lei, não constitua um fator determinante.

Em relação ao terceiro fator da análise do uso justo (a quantidade e a substancialidade do uso em relação à obra protegida), observa-se se a obra foi utilizada em uma proporção razoável para o propósito a se atingir, sendo que o limite considerado “razoável” varia conforme o propósito do uso (como indica a discussão sobre a necessidade do uso no caso Campbell, visto acima) e o efeito sobre o mercado em um dado caso concreto (Sag, 2012, p. 62-63). Esta consideração pode se dar tanto de forma quantitativa como qualitativa: no caso Harper & Row, por exemplo, foram utilizados trechos pequenos de um livro, porém de tamanha relevância no conjunto da obra que seu uso foi considerado excessivo.¹²¹

Por fim, o quarto fator (o impacto do uso sobre o mercado em potencial ou sobre o valor da obra original) trata de uma análise que em muito se assemelha ao segundo elemento da regra dos três passos: o efeito sobre a exploração normal da obra. Mais especificamente, intenta-se determinar se o uso em questão prejudica a exploração comercial usual da obra pelo titular; nota-se que isto inclui não apenas o mercado sobre o original, mas também

¹¹⁸ “Parody needs to mimic an original to make its point, and so has some claim to use the creation of its victim's (or collective victims') imagination, whereas satire can stand on its own two feet and so requires justification for the very act of borrowing” (*Ibid.*, p. 580-81 / 1172-1127).

¹¹⁹ Ver, por exemplo, o caso *Leadsinger, Inc. v. BMG Music Publishing*, 512 F.3d 522 (9th Cir. 2008), p. 531 “Original song lyrics are a work of creative expression, as opposed to an informational work, which is precisely the sort of expression that the copyright law aims to protect.” (Estados Unidos, 2008).

¹²⁰ “Though the right of first publication, like other rights enumerated in § 106, is expressly made subject to the fair use provisions of 107, fair use analysis must always be tailored to the individual case. The nature of the interest at stake is highly relevant to whether a given use is fair. The unpublished nature of a work is a key, though not necessarily determinative, factor tending to negate a defense of fair use. And under ordinary circumstances, the author's right to control the first public appearance of his undissemated expression will outweigh a claim of fair use” (Estados Unidos, 1985, p. 540).

¹²¹ “(iii) Although the verbatim quotes in question were an insubstantial portion of the Ford manuscript, they qualitatively embodied Mr. Ford's distinctive expression, and played a key role in the infringing article” (*Ibid.*, p. 540-541).

sobre a produção de obras derivadas¹²² – aqui entendidas como todas aquelas que normalmente se esperaria que os criadores produzissem ou licenciassem.¹²³

Feitas estas considerações, voltamo-nos para o histórico específico da mineração de textos e dados em relação ao uso justo – que tem sido a forma pela qual os tribunais dos EUA têm lidado com a questão –, tomando como ponto de partida dois casos do início deste século, os quais estabeleceram as bases para a discussão sobre mineração de textos e dados em matéria de Direito Autoral. No primeiro caso, *Kelly v. Arriba Soft* (Estados Unidos, 2003), um fotógrafo processou os proprietários de um mecanismo de busca por violar diretamente os direitos autorais ao acessar fotos em sites, fazer cópias delas nos servidores da Arriba como miniaturas, ou *thumbnails* (as fotos originais seriam então apagadas) e, em seguida, mediante consulta do usuário, mostrar essas miniaturas.¹²⁴ A Corte concluiu que esse uso específico era muito transformador (pois tinha uma finalidade completamente diferente do uso que Kelly fazia de suas fotos) e, portanto, foi considerado uso justo.¹²⁵ O segundo caso foi *Perfect 10 v. Amazon* (Estados Unidos, 2007), também envolvendo o uso de miniaturas em mecanismos de busca. Em uma linha semelhante à de Kelly, a Corte considerou que o uso do réu era justo.¹²⁶

¹²² Para um exemplo da delimitação do quarto fator, ver o caso Campbell: “The fourth fair use factor is “the effect of the use upon the potential market for or value of the copyrighted work.” § 107(4). It requires courts to consider not only the extent of market harm caused by the particular actions of the alleged infringer, but also “whether unrestricted and widespread conduct of the sort engaged in by the defendant ... would result in a substantially adverse impact on the potential market” for the original. Nimmer § 13.05[A][4], p. 13102.61 (footnote omitted); accord, *Harper & Row*, 471 U. S., at 569; Senate Report, p. 65; *Folsom v. Marsh*, 9 F. Cas., at 349. The enquiry “must take account not only of harm to the original but also of harm to the market for derivative works.” *Harper & Row*, *supra*, at 568.” (Estados Unidos, 1994b, p. 590).

¹²³ “The market for potential derivative uses includes only those that creators of original works would in general develop or license others to develop.” (*Ibid.*, 542).

¹²⁴ (“To provide this service, Arriba developed a computer program that “crawls” the web looking for images to index. This crawler downloads full-sized copies of the images onto Arriba’s server. The program then uses these copies to generate smaller, lower-resolution thumbnails of the images. Once the thumbnails are created, the program deletes the full-sized originals from the server. Although a user could copy these thumbnails to his computer or disk, he cannot increase the resolution of the thumbnail; any enlargement would result in a loss of clarity of the image” (Estados Unidos, 2003, p. 815).

¹²⁵ “Kelly asserts that because Arriba reproduced his exact images and added nothing *819 to them, Arriba’s use cannot be transformative. Courts have been reluctant to find fair use when an original work is merely retransmitted in a different medium.¹⁹ Those cases are inapposite, however, because the resulting use of the copyrighted work in those cases was the same as the original use. For instance, reproducing music CDs in computer MP3 format does not change the fact that both formats are used for entertainment purposes. Likewise, reproducing news footage into a different format does not change the ultimate purpose of informing the public about current affairs. (...)

This case involves more than merely a retransmission of Kelly’s images in a different medium. Arriba’s use of the images serves a different function than Kelly’s use—improving access to information on the internet versus artistic expression. Furthermore, it would be unlikely that anyone would use Arriba’s thumbnails for illustrative or aesthetic purposes because enlarging them sacrifices their clarity. Because Arriba’s use is not superseding Kelly’s use but, rather, has created a different purpose for the images, Arriba’s use is transformative” (*Ibid.*, p. 818-819).

¹²⁶ “In this case, Google has put Perfect 10’s thumbnail images (along with millions of other thumbnail images) to a use fundamentally different than the use intended by Perfect 10. In doing so, Google has provided a significant benefit to the public. Weighing this significant transformative use against the unproven use of Google’s thumbnails

No caso *Authors' Guild v. HathiTrust* (Estados Unidos, 2014), autores e grupos de autores processaram a HathiTrust por copiar sistematicamente coleções inteiras de obras em seu repositório, para serem usadas para fins de consulta e preservação. Nesse caso, os membros da HathiTrust eram instituições de ensino/pesquisa e outras organizações sem fins lucrativos, e as obras adicionadas ao banco de dados faziam parte de seu acervo. Além disso, os réus dispunham de um serviço para oferecer obras órfãs, depois de tentar entrar em contato com os respectivos proprietários dos direitos autorais e tentar dar a devida notificação. Nenhuma parte das obras seria mostrada em qualquer pesquisa de usuário – apenas a frequência de aparição de um termo de pesquisa, o que constitui elemento puramente factual. Em vista disso, concluiu-se que essas práticas eram um uso justo, pois (entre outras considerações) eram altamente transformadoras e não representavam nenhum dano previsível ao mercado para as obras incluídas nos repositórios.

Mais emblemático, porém, é o caso *Authors' Guild v. Google* (também conhecido como o caso *Google Books*) (Estados Unidos, 2015). Nele, um grupo de escritores entrou com uma ação contra o Google por digitalizar totalmente suas obras sem permissão, para serem adicionadas ao projeto *Google Books* – o qual envolveu acordos com bibliotecas nos quais obras de seus acervos seriam disponibilizadas para o projeto para serem digitalizadas e transformadas em formato legível para máquinas. A partir daí, o Google criaria um índice com base nesses textos e disponibilizaria trechos das obras para consulta. A cópia também incluiria o armazenamento das imagens digitalizadas dos livros para fins de aprimoramento contínuo do repositório à medida que suas ferramentas tecnológicas fossem aprimoradas. Foi explicitamente reconhecido que esse uso, enquanto mineração de texto e dados, era permitido, com a ressalva de que o projeto tinha restrições sobre como os trechos seriam exibidos e como as obras poderiam ser usadas por terceiros (para evitar a disseminação em larga escala), além de a plataforma não utilizar este projeto para nenhuma atividade com fins lucrativos. No final, o Tribunal de Apelações do 2º Circuito dos EUA decidiu a favor do Google, destacando o *Google Books* como altamente transformador e efetivamente incapaz de permitir o surgimento de um substituto ou um competidor para as obras incluídas no banco de dados do projeto.

Nesse contexto, é possível intuir a atual situação jurídica da mineração de textos e dados perante o Direito Autoral americano. Em primeiro lugar, pode-se argumentar que, muitas vezes, o emprego da mineração de textos e dados requer o uso amplo de obras, quantitativa e

for cell phone downloads, and considering the other fair use factors, all in light of the purpose of copyright, we conclude that Google's use of Perfect 10's thumbnails is a fair use" (Estados Unidos, 2007, p. 1168).

qualitativamente, para ter um desempenho adequado – fato que tem sido aceito como razoável nos tribunais. Além disso, de todos os casos que trataram da coleta e do uso de obras nesse tipo de procedimento, nenhum aparentemente se debruçou sobre as etapas técnicas, mesmo que elas envolvessem a reprodução direta e integral ou a conversão de obras em diferentes formatos. Em vez disso, a consideração central no teste de transformação se concentrou sobre o produto final da mineração: por qual razão ArribaSoft, Google e HathiTrust conduziram suas atividades, como isso se traduziu no resultado final e o quão diferente das obras originais foi este resultado em questão. E, como visto acima, o uso de obras para fins de análise e indexação, para auxiliar nas consultas de pesquisa, é um ponto em comum em todos esses casos - e que foi considerado altamente diferente em termos de propósito, expressão e significado dos livros ou imagens usados.

No entanto, a exposição dos dados a terceiros também é uma questão de preocupação para os tribunais dos EUA: os casos da ArribaSoft, Perfect 10 e Google Books explicitam que, se o usuário secundário, no processo de armazenamento e uso dos dados, não tomar as medidas necessárias para garantir a proteção do conteúdo armazenado - permitindo, assim, que outros usuários se envolvam em atos ilícitos com o material -, isso pesa em desfavor de uma conclusão de uso justo.¹²⁷

Ademais, há atualmente outra questão a ser discutida pelos tribunais estadunidenses: definir o que acontece quando o resultado ou propósito do processo de mineração de textos e dados é aparentemente similar ao uso da obra original, ainda que culmine em uma obra tecnicamente diferente. Como visto anteriormente, sistemas de IA Generativa têm sido utilizados para criar novas obras e têm sido objeto de diversas ações judiciais, ainda sem resolução. A doutrina jurídica norte-americana parece exibir posicionamentos conflitantes nesse sentido: por exemplo, Sobel (2017), como visto anteriormente, já observava o aumento desse tipo de uso e questionava o alcance da doutrina sobre uso não-expressivo, alertando para uma inversão na qual grandes usuários (como Big Techs) estariam se beneficiando do uso justo, obtendo valor econômico para além do valor social que a doutrina pretende incentivar.¹²⁸ Outros autores

¹²⁷“If, in the course of making an arguable fair use of a copyrighted work, a secondary user unreasonably exposed the rights holder to destruction of the value of the copyright resulting from the public's opportunity to employ the secondary use as a substitute for purchase of the original (even though this was not the intent of the secondary user), this might well furnish a substantial rebuttal to the secondary user's claim of fair use. For this reason, the Arriba Soft and Perfect 10 courts, in upholding the secondary user's claim of fair use, observed that thumbnail images, which transformatively provided an Internet pathway to the original images, were of sufficiently low resolution that they were not usable as effective substitutes for the originals. Arriba Soft, 336 F.3d 811 at 819; Perfect 10, 508 F.3d at 1165”(Estados Unidos, 2015, p. 227).

¹²⁸ “In today's platform economy, value emerges not from the ownership of intellectual property rights in data, but from the ability to make licensed use of large amounts of data. Big Users are hegemonic, not Big Content. This

mantém não apenas a noção de que certos usos, como cópias incidentais para análise, permanecem não-expressivos (e, portanto, lícitos), como também que a legalidade deste uso não está no input, mas no propósito e efeitos da utilização no resultado final desses processos. Sag (2019), por exemplo, é enfático quanto à importância da licitude do uso não-expressivo pelo resultado final: se o produto final de um processo de coleta e análise é meramente a geração de informação a respeito de uma expressão sem usá-la explicitamente, então tal uso não poderia ter efeito substitutivo sobre uma obra.¹²⁹ Já Carroll (2020) não apenas expressa preocupação com os possíveis efeitos de uma inversão da lógica “pró-pequeno usuário” habitual no modelo norte-americano,¹³⁰ mas argumenta que, ao menos no que diz respeito à mineração de textos e dados para fins de análise, certos usos não-expressivos e mesmo outros, mais expressivos, necessitam ser resguardados pelo uso justo.¹³¹

Na prática, responder a essa questão passa por uma série de problemáticas: além de definir se estilos artísticos constituem expressão ou se estão mais próximos de ideias (algo já estabelecido no Brasil, conforme discutido no Capítulo 2), há também a necessidade de estabelecer até que ponto o uso desses programas concorre com as obras utilizadas para treiná-los: por um lado, o quarto fator da análise de uso justo pondera a respeito dos efeitos concorrenciais sobre uma obra em particular, e não sobre a produção do seu autor. Por outro lado, o mesmo fator também analisa o efeito sobre o *valor* daquela obra. Nesse sentido, ao fornecer uma IA capaz de gerar uma obra com estilo parecido com o de um artista em particular, estaria alguém provocando uma redução na procura pelo licenciamento das obras criadas por este artista – e, portanto, uma desvalorização de suas obras? E esta hipótese sequer se enquadraria no espírito

reshuffling of owners and users demands a corresponding readjustment in intuitions about fair use (...) Commercial machine learning, trained on expressive media, promises tremendous social value. But it is not the sort of value that fair use exists to foster. Unlike the benefits realized by, say, scholarship, the value of advanced machine learning services is internalized by the large firms that furnish those services.” (Sobel, 2017, p. 88-89).

¹²⁹ “As discussed above, copyright law is fundamentally concerned with the communication of original expression to the public and not with the mechanical act of copying per se (...). differently: non-expressive uses pose no threat of expressive substitution. Non-expressive uses generate information about a work, that information may be useful, it may be valuable, it may even affect the demand for the underlying work, but metadata about a work does not in any way fulfill the public’s demand for the author’s original expression..” (Sag, 2019, p. 17-18).

¹³⁰ “It is also worth noting that the fair use analysis that justifies TDM research on the data also is closely related to the issue of whether using such works as training data for machine learning systems or other forms of artificial intelligence is a fair use (...) (R)eshaping the relation between transformative uses and the market analysis in fair use risks creating serious collateral damage for other fair uses.” (Carroll, 2020, p. 938).

¹³¹ “Much of this literature concludes that fair use permits “non-expressive” or “non-consumptive” uses of copyrighted works (...). In the TDM context, it appears that this scholarly analysis has focused on whether making temporary copies for the purposes of extracting noncopyrightable information from copyrighted works is fair use. I agree that fair use permits making such copies, but I further argue that fair use also is needed to justify assembling and keeping the dataset used in TDM research (...). Fair use justifies making and keeping a database of copyrighted works necessary to enable computational analysis and to reproduce the results of such research. The “non-expressive” label is less helpful when applied to this use, but the case law provides strong support for this aspect of computational research to also be fair use.” (*Ibid.*, p. 937. Adaptações feitas pelo autor).

do que o quarto fator de análise prevê? Esta é uma dúvida que as cortes norte-americanas talvez precisem enfrentar em breve.

De todo modo, o modelo de cláusula aberta exemplificado no uso justo fornece uma possível via de regulação, que apresenta algumas vantagens. A primeira é que este modelo trata de todos os direitos exclusivos, concentrando-se no *uso* de forma geral, e não apenas na reprodução ou adaptação de obras.

Outra vantagem é que tal modelo não impõe óbice automático à observação de finalidade econômica, embora isto constitua um fator-chave na análise da finalidade e do caráter do uso secundário. Essencialmente, ao invés de optar por um modelo binário (comercial = ruim e não-comercial = bom), as cortes medem a relação entre o lucro econômico privado e o interesse público: o primeiro pesa contra o uso, enquanto o segundo o coloca sob uma luz mais positiva.¹³² Na prática, isto significa que o uso com fins lucrativos, por si só, não exclui automaticamente uma conclusão de uso justo – afinal, em *Campbell*, observa-se que várias atividades explicitamente mencionadas como uso justo na seção 107 são frequentemente conduzidas com fins lucrativos. Da mesma forma, um uso sem fins lucrativos ainda pode ser considerado infrigente se substituir a obra original.¹³³

Já o acesso lícito não é um fator explícito em um julgamento de uso justo, ainda que possa ser indiretamente intuído: por exemplo, no caso *Harper & Row v. Nation Enterprises*

¹³² “Consistent with these principles, courts will not sustain a claimed defense of fair use when the secondary use can fairly be characterized as a form of “commercial exploitation,” i.e., when the copier directly and exclusively acquires conspicuous financial rewards from its use of the copyrighted material. See *Harper & Row*, 471 U.S. at 562–63, 105 S.Ct. at 2231–32; *Twin Peaks*, 996 F.2d at 1375; *Rogers*, 960 F.2d at 309; *Iowa State University Research Foundation, Inc. v. American Broadcasting Companies, Inc.*, 621 F.2d 57, 61 (2d Cir.1980); *Meeropol v. Nizer*, 560 F.2d 1061, 1069 (2d Cir.1977) (examining whether use was “predominantly for commercial exploitation”), cert. denied, 434 U.S. 1013, 98 S.Ct. 727, 54 L.Ed.2d 756 (1978). Conversely, courts are more willing to find a secondary use fair when it produces a value that benefits the broader public interest. See *Twin Peaks*, 996 F.2d at 1375; *Sega Enterprises*, 977 F.2d at 1523; *Rosemont Enterprises, Inc. v. Random House, Inc.*, 366 F.2d 303, 307–09 (2d Cir.1966), cert. denied, 385 U.S. 1009, 87 S.Ct. 714, 17 L.Ed.2d 546 (1967). The greater the private economic rewards reaped by the secondary user (to the exclusion of broader public benefits), the more likely the first factor will favor the copyright holder and the less likely the use will be considered fair” (Estados Unidos, 1994a, p. 922).

¹³³ “The language of the statute makes clear that the commercial or nonprofit educational purpose of a work is only one element of the first factor enquiry into its purpose and character. Section 107(1) uses the term “including” to begin the dependent clause referring to commercial use, and the main clause speaks of a broader investigation into “purpose and character.”(...) [T]he mere fact that a use is educational and not for profit does not insulate it from a finding of infringement, any more than the commercial character of a use bars a finding of fairness. If, indeed, commerciality carried presumptive force against a finding of fairness, the presumption would swallow nearly all of the illustrative uses listed in the preamble paragraph of § 107, including news reporting, comment, criticism, teaching, scholarship, and research, since these activities “are generally conducted for profit in this country.” *Harper & Row*, supra, at 592, 105 S.Ct., at 2246 (Brennan, J., dissenting). Congress could not have intended such a rule, which certainly is not inferable from the common-law cases, arising as they did from the world of letters in which Samuel Johnson could pronounce that “[n]o man but a blockhead ever wrote, except for money.” 3 *Boswell's Life of Johnson* 19 (G. Hill ed. 1934)”(Estados Unidos, 1994b, p. 584/1174).

(Estados Unidos, 1985), entendeu-se que o uso não-autorizado dos trechos de um livro ainda não publicado feria o direito do titular sobre a primeira publicação da obra. Dito isto, este foi apenas um dos elementos a serem considerados no caso, não constituindo um fator determinante de forma isolada – de forma que há posicionamentos, tanto na doutrina (por exemplo, ver Samuelson e Asay, 2018) como na jurisprudência (a exemplo do caso Campbell, onde a autorização para uso havia sido expressamente negada) (Estados Unidos, 1994b), no sentido de que mesmo o uso explicitamente não-licenciado., por si só, não preclui automaticamente a observância do uso justo no caso concreto.

De forma mais ampla, a cláusula aberta também implica, essencialmente, em fazer com que a definição dos usos livres dentro do regime de limitações e exceções deixe de ser uma prerrogativa exclusiva do Poder Legislativo, concedendo ao Poder Judiciário a prerrogativa de expandir e atualizar o regime de limitações e exceções de Direito Autoral sem a necessidade de intervenção via processo legislativo¹³⁴ – criando, portanto, um caminho alternativo para a criação de novos usos permitidos sem a necessidade de passar pelas etapas do processo legislativo e, ao mesmo tempo, permitindo uma participação social mais direta na construção das L&E, uma vez que usos justos criados por este sistema resultariam de demandas diretamente manifestadas pelo jurisdicionado, ao invés de discussões conduzidas por representantes eleitos.

Tal liberdade, combinada ao caráter principiológico dos fatores, também confere ao modelo de cláusula aberta o que talvez seja sua principal vantagem: dá ao regime de limitações e exceções (e ao Direito de Autor, de forma mais geral) uma maior flexibilidade para se ajustar a novas demandas do tempo, uma vez que as cortes podem permitir e delimitar hipóteses oriundas de novas circunstâncias sociais e tecnológicas, não estando limitados aos usos livres já previstos em lei.¹³⁵

¹³⁴ O que não significa que não haja preocupações em relação a questões como inovação, tal como se poderia ver no Poder Executivo ou Legislativo. Pelo contrário, não é incomum que as cortes, ao julgarem casos de uso justo, manifestem, direta ou indiretamente, preocupações com os efeitos de suas decisões sobre mercados e tecnologias. Vide, por exemplo, a decisão do caso *Sony Corp. of America v. Universal City Studios, Inc.* (o caso Sony Betamax) nos EUA, envolvendo o uso de aparelhos de videocassete para gravar programas televisivos, no qual a Suprema Corte afirma que “Se há milhões de proprietários de VTRs que fazem cópias de eventos esportivos televisionados, transmissões religiosas e programas educacionais como *Mister Rogers' Neighborhood*, e se os proprietários desses programas aceitam essa prática, o empreendimento de fornecer o equipamento que viabiliza essa cópia não deve ser sufocado simplesmente porque o equipamento é usado por alguns indivíduos para fazer reproduções não autorizadas das obras dos réus.” (Estados Unidos, 1984).

¹³⁵ Conforme afirmado no próprio Congresso dos EUA: “The bill endorses the purpose and general scope of the judicial doctrine of fair use, but there is no disposition to freeze the doctrine in the statute, especially during a period of rapid technological change. Beyond a very broad statutory explanation of what fair use is and some of the criteria applicable to it, the courts must be free to adapt the doctrine to particular situations on a case-by-case basis...” (Estados Unidos, 1976b, p. 65).

Entretanto, é importante considerar que uma cláusula aberta trata especificamente do estabelecimento de usos livres, mas não implica em um posicionamento inerente em relação, por exemplo, à possibilidade de estipulação de cláusula contratual na qual se renuncie do direito de lançar mão desse recurso pela via judicial – questão essa que, independentemente da existência desta cláusula, deve ser regulada à parte, seja pelo próprio Judiciário ou pela via legal. Com efeito, no caso específico dos EUA, há jurisprudência no sentido de que o *Copyright Act* não impediria a possibilidade de renúncia contratual do uso de obras para fins já considerados usos justos no país.¹³⁶

De forma similar, a possibilidade de contornar medidas tecnológicas de proteção não é algo automaticamente garantido por este modelo – dependendo, novamente, do entendimento da jurisdição local. No caso dos Estados Unidos, a jurisprudência considera que a Seção 1201 do *Digital Millenium Copyright Act* (o DMCA), que proíbe a burla desses dispositivos,¹³⁷ não

¹³⁶ Para um exemplo, ver *Bowers v. Baystate Techs., Inc.*, 320 F.3d 1317, 1325–26 (Fed. Cir. 2003) (“Indeed, the Supreme Court recently noted “[i]t goes without saying that a contract cannot bind a nonparty.” *EEOC v. Waffle House, Inc.*, 534 U.S. 279, 122 S.Ct. 754, 764, 151 L.Ed.2d 755 (2002). This court believes that the First Circuit would follow the reasoning of ProCD and the majority of other courts to consider this issue. This court, therefore, holds that the Copyright Act does not preempt Mr. Bowers' contract claims (...). (P)ivate parties are free to contractually forego the limited ability to reverse engineer a software product under the exemptions of the Copyright Act”); *Davidson & Assocs. v. Jung*, 422 F.3d 630, 638–39 (8th Cir. 2005) (“Appellants, relying upon *Vault v. Quaid Software Ltd.*, 847 F.2d 255, 268–70 (5th Cir.1988), argue that the federal Copyright Act preempts Blizzard's state law breach-of-contract claims. We disagree (...). Unlike in *Vault*, the state law at issue here neither conflicts with the interoperability exception under 17 U.S.C. § 1201(f) nor restricts rights given under federal law. Appellants contractually accepted restrictions on their ability to reverse engineer by their agreement to the terms of the TOU and EULA” (Estados Unidos).

¹³⁷ 17 U.S. Code § 1201 - Circumvention of copyright protection systems

(a) Violations Regarding Circumvention of Technological Measures.—

(1)

(A) No person shall circumvent a technological measure that effectively controls access to a work protected under this title. The prohibition contained in the preceding sentence shall take effect at the end of the 2-year period beginning on the date of the enactment of this chapter.

(B) The prohibition contained in subparagraph (A) shall not apply to persons who are users of a copyrighted work which is in a particular class of works, if such persons are, or are likely to be in the succeeding 3-year period, adversely affected by virtue of such prohibition in their ability to make noninfringing uses of that particular class of works under this title, as determined under subparagraph (C).

(...)

(E) Neither the exception under subparagraph (B) from the applicability of the prohibition contained in subparagraph (A), nor any determination made in a rulemaking conducted under subparagraph (C), may be used as a defense in any action to enforce any provision of this title other than this paragraph.

(2) No person shall manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that—

(A) is primarily designed or produced for the purpose of circumventing a technological measure that effectively controls access to a work protected under this title;

(B) has only limited commercially significant purpose or use other than to circumvent a technological measure that effectively controls access to a work protected under this title; or

(C) is marketed by that person or another acting in concert with that person with that person's knowledge for use in circumventing a technological measure that effectively controls access to a work protected under this title.

(...)

(b) Additional Violations.—

seria alcançada pelo *fair use* – que, portanto, não constitui escusa legalmente aceitável neste caso.¹³⁸ Curiosamente, entretanto, o próprio DMCA, estabelece um rol fechado de limitações e exceções, estipulando casos nos quais esta burla é considerada lícita. Tal lista inclui limitações existentes no próprio texto da Seção,¹³⁹ bem como critérios¹⁴⁰ para o estabelecimento de uma

(1) No person shall manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that—

(A) is primarily designed or produced for the purpose of circumventing protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof;

(B) has only limited commercially significant purpose or use other than to circumvent protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof; or

(C) is marketed by that person or another acting in concert with that person with that person’s knowledge for use in circumventing protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof.

(...)

(c) Other Rights, Etc., Not Affected.—

(1) Nothing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title.

(2) Nothing in this section shall enlarge or diminish vicarious or contributory liability for copyright infringement in connection with any technology, product, service, device, component, or part thereof.

(3) Nothing in this section shall require that the design of, or design and selection of parts and components for, a consumer electronics, telecommunications, or computing product provide for a response to any particular technological measure, so long as such part or component, or the product in which such part or component is integrated, does not otherwise fall within the prohibitions of subsection (a)(2) or (b)(1).

(4) Nothing in this section shall enlarge or diminish any rights of free speech or the press for activities using consumer electronics, telecommunications, or computing products.” (Estados Unidos, 1998).

¹³⁸ “The Court begins its statutory analysis, as it must, with the language of the statute. Section 107 of the Copyright Act provides in critical part that certain uses of copyrighted works that otherwise would be wrongful are “not . . . infringement[s] of copyright.” Defendants, however, are not here sued for copyright infringement. They are sued for offering and providing technology designed to circumvent technological measures that control access to copyrighted works and otherwise violating Section 1201(a)(2) of the Act. If Congress had meant the fair use defense to apply to such actions, it would have said so. Indeed, as the legislative history demonstrates, the decision not to make fair use a defense to a claim under Section 1201(a) was quite deliberate.” (Estados Unidos, 2000).

¹³⁹ Isto inclui limitações e exceções para bibliotecas e arquivos abertos ao público, atividades governamentais e de inteligência, engenharia reversa, pesquisas com criptografia, controle de acesso a menores de idade, proteção de dados pessoais e testes de segurança. Tais casos estão previstos na Seção 1201, (d) a (j) do Copyright Act (Estados Unidos, 1998).

¹⁴⁰ “(C) During the 2-year period described in subparagraph (A), and during each succeeding 3-year period, the Librarian of Congress, upon the recommendation of the Register of Copyrights, who shall consult with the Assistant Secretary for Communications and Information of the Department of Commerce and report and comment on his or her views in making such recommendation, shall make the determination in a rulemaking proceeding for purposes of subparagraph (B) of whether persons who are users of a copyrighted work are, or are likely to be in the succeeding 3-year period, adversely affected by the prohibition under subparagraph (A) in their ability to make noninfringing uses under this title of a particular class of copyrighted works. In conducting such rulemaking, the Librarian shall examine—

(i) the availability for use of copyrighted works;

(ii) the availability for use of works for nonprofit archival, preservation, and educational purposes;

(iii) the impact that the prohibition on the circumvention of technological measures applied to copyrighted works has on criticism, comment, news reporting, teaching, scholarship, or research;

(iv) the effect of circumvention of technological measures on the market for or value of copyrighted works; and

(v) such other factors as the Librarian considers appropriate.

(D) The Librarian shall publish any class of copyrighted works for which the Librarian has determined, pursuant to the rulemaking conducted under subparagraph (C), that noninfringing uses by persons who are users of a copyrighted work are, or are likely to be, adversely affected, and the prohibition contained in subparagraph (A) shall not apply to such users with respect to such class of works for the ensuing 3-year period.”

lista fechada (porém mais extensiva) sob a curadoria da Biblioteca do Congresso dos EUA, a qual passa por atualizações a cada três anos – nas quais usos já previstos podem ser renovados e novos usos livres podem ser propostos e incluídos (Estados Unidos, 2025). A última versão da lista foi publicada em outubro de 2024, renovando praticamente todos os usos livres já previstos na versão anterior (Estados Unidos, 2024b).¹⁴¹

Aqui, cumpre reiterar, contudo, que o não-alcançamento do *fair use* a dispositivos contratuais e a regulações anti-burla constitui uma característica do Direito Autoral dos EUA, e não do modelo de cláusula aberta em si. Nesse sentido, nada impediria que outra jurisdição com este ou outro tipo de regulação tenha outro entendimento quanto aos pontos acima. Inclusive, como se verá no Capítulo 5, o Brasil já dispõe de uma cláusula aberta, pautada na função social da propriedade e na observância a outros direitos fundamentais, que serviriam como elementos norteadores para dispor sobre o status das limitações e exceções perante os contratos e as medidas tecnológicas de proteção – e, considerando-se a doutrina do abuso de direito (Souza, A., 2009), potencialmente impedir que estes mecanismos interfiram com o uso e usufruto das limitações e exceções garantidas seja por lei, seja pelo Judiciário.

De qualquer modo, a cláusula aberta se coloca como ferramenta poderosa à disposição do Judiciário como via alternativa ao Legislativo e instrumento de flexibilização e abertura no sistema de Direito Autoral, desde que cumprida uma condição: o acionamento das cortes para que possam lançar mão deste recurso. Mais especificamente, este tipo de cláusula, a exemplo da doutrina do *fair use*, apenas estipula como o Judiciário pode criar esses casos, mas não implica, automaticamente, na existência de qualquer limitação específica a menos que esta seja construída através de casos concretos. Isto implica em dizer que, para que um “certo caso especial” (nos termos da Convenção de Berna) seja estabelecido por meio da cláusula aberta judicial, alguém necessita acionar os tribunais por meio de um processo litigioso.

Nesse sentido, não é à toa que Lessig (2004, p. 168) já se referiu à doutrina do *fair use* como “o direito de contratar um advogado”: o jurista comenta que diversos advogados consideravam o licenciamento de obras para fins de arquivamento como um “uso justo” – e, portanto, permitido dentro do Direito Autoral local – sem atentar, contudo, para o custo de fazer valer a licitude desse uso: litígios, de uma forma geral, podem envolver o dispêndio de tempo e recursos financeiros que muitos profissionais ou instituições que lidam com pesquisa podem não ter

¹⁴¹ “The Register received petitions to renew all but one of the exemptions adopted pursuant to the eighth triennial rulemaking, and recommends renewal of all exemptions for which petitions were filed”. A exceção que não recebeu petição para renovação diz respeito à burla de dispositivos de proteção em jogos eletrônicos feita por usuários portadores de deficiência para o uso de ferramentas de acessibilidade (Estados Unidos, 2024b).

condições (ou disposição) de assumir.¹⁴² De forma similar, é possível presumir que, se uma empresa restringe injustificadamente o acesso à API de um determinado programa ou aplicativo, qualquer tentativa de se buscar uma solução judicial para o problema também resultará em despesas que um novo entrante pode não estar apto ou disposto a pagar. Nesse sentido, nota-se que, dentre todos os casos envolvendo mineração de textos e dados mencionados nesta seção, as partes envolvidas são, geralmente, compostas por grandes empresas. O próprio HathiTrust constitui uma exceção que confirma a regra, sendo composto por diversas instituições dedicadas ao ensino e pesquisa.

Por fim, há ainda uma questão que resulta da própria configuração das fontes do Direito, e que diferencia usos livres oriundos de decisões judiciais daqueles que emanam do legislador: a própria generalidade e abstração dos diferentes tipos de normas jurídicas. Mais especificamente, Bobbio (2010, p. 177) aponta que a maior parte das leis são, de fato, gerais e abstratas, dirigindo-se a todos os cidadãos e desenhadas para abranger uma amplitude razoável de casos. Já sentenças judiciais são o oposto: individuais e concretas, estando circunscritas às pessoas envolvidas em um determinado caso e abrangendo apenas as circunstâncias factuais específicas deste.¹⁴³ Assim, por esta lógica, embora certos elementos de uma decisão judicial possam se repetir em situações futuras, o que foi decidido em um litígio não pode ser automaticamente extrapolado para todo um grupo de casos: mudanças significativas nos fatos exigirão novos julgamentos e novas ponderações.

Assim, a mineração de textos e dados é oficialmente considerada uso justo nos Estados Unidos, mas apenas em situações que se assemelhem aos casos já julgados. Por exemplo: os casos HathiTrust e Google Books ambos lidaram com o uso de material protegido para mineração de textos, mas não implicaram na permissibilidade de qualquer uso para mineração de

¹⁴² “Mas o uso justo na América simplesmente significa o direito de contratar um advogado para defender seu direito de criar. E como os advogados adoram se esquecer, nosso sistema para defender causas como a de uso justo é impressionantemente ruim — em praticamente todos os contextos, mas principalmente aqui. Ele custa muito, demora muito, e normalmente o que ele delibera tem pouca conexão com a justiça real no caso. O sistema legal pode ser tolerável para os muito ricos. Para os demais, é um embaraço em uma tradição que sempre se orgulhou da importância das leis.” (Lessig, 2004, p. 168).

¹⁴³ Explica-se: Bobbio (2010, p. 174-177) aponta que, de forma geral, a doutrina do Direito simplesmente aponta que normas jurídicas seriam gerais e abstratas, quando na verdade existiriam em um duplo espectro. Um dos eixos se refere à universalidade dos indivíduos a quem a norma se dirige (sendo gerais as que se destinam a uma classe de pessoas e individuais, as que se voltam para indivíduos específicos), ocupando-se o outro eixo da universalidade das ações (sendo abstratas as normas que regulam um tipo ou classe de ações e concretas, as que descrevem uma ação singular). Assim, normas jurídicas existiriam em quatro tipos: gerais e abstratas (a maioria das leis, como a penal), gerais e concretas (uma lei que declare uma mobilização geral para um dado evento, como no caso da Lei da Copa), individuais e abstratas (como uma lei que atribui a uma pessoa uma determinada função pública e todos os atos oficiais dali decorrentes) e individuais e concretas (a exemplo de decisões judiciais, que se dirigem a indivíduos específicos e dispõem sobre atos específicos).

textos e dados: apenas aqueles em que foram utilizados textos legalmente acessados, com condições estritas de segurança, oferecendo-se apenas pequenos trechos para visualização (ou apenas metadados), sem fins comerciais, entre outras exigências. Isto pode levar à impressão de que, no geral, usos livres seriam apenas tão específicos quanto os vistos em listas fechadas de limitações e exceções, mas não apenas há regulações mais amplas a respeito (como se verá a seguir) como não se sabe se, e até que ponto, estariam permitidos os processos de mineração de textos e dados em IA, ou mesmo em outros casos que façam uso de músicas, pinturas, vídeos ou outras mídias, ou cujo acesso ao material tenha se originado de material não-licenciado. Se não há precedente exato, não é possível definir, *ex ante*, o que (e sob quais condições) está efetivamente permitido.

Dessa maneira, o modelo de cláusula aberta, de forma mais ampla, pode ser entendido não como um substituto para a via legislativa, mas como um mecanismo concomitante ao processo legislativo para a estipulação e aplicação dos usos livres. Nesse sentido, fortalece as limitações e exceções não só mediante o sopesamento de direitos fundamentais, mas também ao garantir que mais atores participem da sua criação. Além disso, oferece uma maior adaptabilidade do sistema de Direito Autoral às mudanças tecnológicas e sociais, servindo como uma alternativa potencialmente mais rápida do que processos de reforma legislativa. Por outro lado, isto é compensado por uma maior restritividade no escopo dos usos livres resultantes desse sistema em relação a uma limitação fechada; ademais, tal cláusula, para ser empregada em seu potencial ideal, necessita da mobilização de agentes aptos e dispostos a arcar com os custos monetários, temporais e mentais de uma disputa judicial.

Nesse sentido, Yu (2019) menciona que é comum muitos países, ao emularem o modelo de cláusula aberta – mais especificamente, a doutrina do uso justo –, adotarem modelos mistos (com cláusulas abertas acompanhadas de listas fechadas, exemplificativas). É o caso, por exemplo, do Equador, que possui uma cláusula aberta além de uma lista extensa de usos livres (Código Ingenios, art. 211)¹⁴⁴ e de Singapura, que inclusive possui uma limitação específica para mineração de textos e dados.¹⁴⁵ Os próprios Estados Unidos também se enquadram

¹⁴⁴ Os critérios de análise são (i) o objetivo e natureza do uso, (ii) a natureza da obra, (iii) a quantidade e a importância da parte usada em relação ao seu conjunto, (iv) o efeito do uso do valor de mercado da obra, e (v) o exercício de outros direitos fundamentais. (Equador, 2016).

¹⁴⁵ Copying or communicating for computational data analysis

244.—(1) If the conditions in subsection (2) are met, it is a permitted use for a person (X) to make a copy of any of the following material:

(a) a work;
(b) a recording of a protected performance.

neste caso: além da lista de exceções ao DMCA expostas acima, o *Copyright Act* possui um rol de L&E logo após a cláusula de fair use, em suas Seções 108 a 112, 117, 119 e 121 a 122.

Ao mesmo tempo, outro demonstrativo do caráter híbrido desse sistema é o fato de constatações de uso justo não obstarem o exercício, no Congresso, de esforços regulatórios em um sentido diferente: a explosão da IA Generativa já tem motivado movimentos no Congresso norte-americano para impor limites ao que pode ser utilizado na mineração de textos e dados. O exemplo mais recente é a proposição do *COPIED Act - Content Origin Protection and Integrity from Edited and Deepfaked Media Act of 2024*, de 11 de julho de 2024 (Estados Unidos, 2024a),¹⁴⁶ que tem como propósito justamente a regulação do uso de obras protegidas por sistemas de IA. Dentre as disposições propostas, prevê-se o estabelecimento de diretrizes para a identificação e uso de material protegido (aqui listado como “conteúdo coberto”), bem como uma proibição para o uso não-autorizado de obras protegidas no treinamento de IA.¹⁴⁷ Trata-se

(2) The conditions are —
 (a) the copy is made for the purpose of —
 (i) computational data analysis; or
 (ii) preparing the work or recording for computational data analysis;
 (b) X does not use the copy for any other purpose;
 (c) X does not supply (whether by communication or otherwise) the copy to any person other than for the purpose of —
 (i) verifying the results of the computational data analysis carried out by X; or
 (ii) collaborative research or study relating to the purpose of the computational data analysis carried out by X;
 (d) X has lawful access to the material (called in this section the first copy) from which the copy is made; and
 (e) one of the following conditions is met:
 (i) the first copy is not an infringing copy; (ii) the first copy is an infringing copy but —
 (A) X does not know this; and
 (B) if the first copy is obtained from a flagrantly infringing online location (whether or not the location is subject to an access disabling order under section 325) — X does not know and could not reasonably have known that;
 (iii) the first copy is an infringing copy but —
 (A) the use of infringing copies is necessary for a prescribed purpose; and
 (B) X does not use the copy to carry out computational data analysis for any other purpose.” (Singapura, 2021).

¹⁴⁶ O texto do projeto se encontra disponível em: <https://www.congress.gov/bill/118th-congress/senate-bill/4674/text>. Acesso em 30 ago. 2024.

¹⁴⁷ “SEC. 4. Facilitation of development of standards for content provenance information and detection of synthetic content and synthetically-modified content.

(a) In general.—The Under Secretary shall establish a public-private partnership to facilitate the development of standards regarding content provenance information technologies and the detection of synthetic content and synthetically-modified content, including with respect to the following:

(1) Facilitating the development of guidelines and voluntary, consensus-based standards and best practices for watermarking, content provenance information, synthetic content and synthetically-modified content detection, including for images, audio, video, text, and multimodal content, the use of data to train artificial intelligence systems, and such other matters relating to transparency of synthetic media as the Under Secretary considers appropriate.

(...)

SEC. 6. Requirements for content provenance information; prohibited acts.

(...)

(c) Prohibition on non-Consensual use of covered content that has attached or associated content provenance information.—It shall be unlawful for any person, for a commercial purpose, to knowingly use any covered content that has content provenance information that is attached to or associated with such covered content or covered

de uma iniciativa que já conta com o apoio de entidades de representação da indústria cultural, como o SAG-AFTRA no ramo audiovisual e a RIAA, Universal, Sony e Warner, no setor fonográfico.¹⁴⁸

No mais, como se observará posteriormente neste trabalho, argumentamos que o Brasil também adota um modelo híbrido, com um rol exemplificativo de usos livres acompanhado de uma regra geral, estabelecida no Judiciário, para o estabelecimento de permissões adicionais. Contudo, conforme já mencionado, a legislação brasileira ainda não conta com decisões, ao menos na esfera federal, que abordem o tema da coleta e análise de dados no Direito Autoral. Tampouco há, por ora, uma limitação específica sobre o assunto – embora o PL 2.338/2023 seja o mais próximo que o país chegou nesse sentido. Por esta razão, convém voltar-se para outra forma de regular atividades de mineração de textos e dados no Direito Autoral: a criação de cláusulas específicas, em lei, sobre o assunto.

3.2. Os modelos de limitação estatutária para mineração de textos e dados

Embora o uso justo estadunidense tenha lançado, ainda no início dos anos 2000, as bases jurídicas para proteger certos usos de mineração de textos e dados perante o Direito Autoral, o sistema de cláusula aberta não foi o primeiro a produzir uma limitação explícita para a mineração de textos e dados. Tampouco parece ser o modo mais popular de se fazê-lo: o mapa de Flynn e Palmedo (2022), exposto anteriormente, indica que apenas uma pequena parcela de países recebeu a classificação verde – categoria que conta com a participação total, mas não exclusiva, das jurisdições com cláusulas abertas observados no estudo. A maioria dos países que possui alguma previsão legal nesse sentido opta, portanto, por um caminho distinto: a inclusão de processo de mineração de textos e dados dentro do rol de usos livres previstos explicitamente em lei.

A primeira vista, trata-se de um modelo menos flexível que o de cláusula aberta: ao contrário dos EUA, onde houve (e haverá) diversas oportunidades de examinar usos específicos da coleta e análise de dados mediante acionamento do Poder Judiciário, todas as limitações e exceções que serão vistas nesta seção foram produto de extensa deliberação legislativa, com atualizações sendo a exceção, não a regra. Por outro lado, como já dito, este modelo apresenta

content from which the person knows or should know that content provenance information has been removed or separated in violation of subsection (b), in order to train a system that uses artificial intelligence or an algorithm or to generate synthetic content or synthetically-modified content unless such person obtains the express, informed consent of the person who owns the covered content, and complies with any terms of use pertaining to the use of such content, including terms regarding compensation for such use, as required by the owner of copyright in such content.” (Estados Unidos, 2024).

¹⁴⁸ Ver UBC (2024).

uma vantagem: o fato de não estar limitado a circunstâncias fáticas específicas, podendo cobrir um espectro muito mais amplo de casos. Isto também impõe uma responsabilidade singular: evitar uma sobreinclusão (quando a norma alcança situações para as quais não foi planejada) e, ao mesmo tempo, uma subinclusão (quando não se atingem todos os casos em que a aplicação da lei seria desejada) (Schauer, 1991).

Esta seção, serão observadas algumas tentativas de enfrentar esses dilemas, dentro de um contexto que se destaca por uma visão centrada no estímulo ao compartilhamento de dados como um fator contemporâneo de incentivo à inovação e ao desenvolvimento econômico – tema abordado de forma mais específica em Alvarenga (2019) no tocante à atividade empresarial. Aqui, propõe-se olhar para os modelos regulatórios seguintes de forma um pouco diferente, não enquanto ferramentas à serviço de companhias, mas sim da ciência. Ao mesmo tempo, a documentação aqui exposta também sugere que os casos aqui expostos, de maneira geral, ressaltam a noção de uma academia cada vez mais inseparável de considerações econômicas, dentro do atual paradigma das políticas de inovação – conforme discutiremos no capítulo seguinte.

3.2.1. Usos livres gerais para pesquisa

Antes de se avançar para regulações específicas para a mineração automatizada de textos e dados, é pertinente destacar que estas técnicas constituem apenas um dentre diversos métodos de coleta e análise no ambiente acadêmico. Assim, pode-se deduzir que, ao invés de um dispositivo legal específico para a mineração de textos e dados, é possível legalizar estas práticas de forma mais indireta: por meio de hipóteses legais que permitem o uso de obras protegidas para fins de pesquisa de forma mais geral. Esta opção teria a vantagem de permitir práticas de investigação científica para além da coleta e análise computacional, atendendo a um número maior de pesquisadores. Além disso, no caso de legislações que já possuem disposições legais para a pesquisa, basta simplesmente interpretar o uso da mineração de textos e dados como implicitamente amparado pela lei quando empregado na investigação científica, não exigindo a criação de novas normas.

Nesse sentido, notam-se, ao redor do mundo, diferentes opções regulatórias neste sentido: é possível permitir usos com propósito de pesquisa de uma forma mais ampla, (a exemplo da Nigéria, que simplesmente estabelece a pesquisa não-comercial entre as hipóteses sobre as quais o fair dealing se aplicaria),¹⁴⁹ ou estabelecer um dispositivo próprio, detalhando as

¹⁴⁹ “Art 51. The right conferred in respect of a work by section 5 of this Act does not include the right to control—
(a) the doing of any of the acts mentioned in the said section 5 by way of **fair dealing for purposes of research**, private use, criticism or review or the reporting of current events, subject to the condition that, if the use is public,

circunstâncias específicas para sua aplicação (como no caso do Reino Unido, que prevê uma hipótese legal do gênero para além da sua limitação mais específica para coleta e análise de dados no art. 29A).¹⁵⁰

Na América Latina, a legislação de Direito Autoral equatoriana aparece novamente como um exemplo relevante nesse sentido: sua lista extensa de usos livres também conta com hipóteses para contextos educacionais e de pesquisa. Para fins educativos, o *Código Ingenios* permite a comunicação (art. 212 num. 17 e 28), a reprodução (art. 212 num. 14 e 15) e a seleção de leituras (art. 212 num. 1 inc. 2). Existe também uma limitação relativa à educação online (art. 212 num. 15), e uma de engenharia reversa de *software* para fins de pesquisa e educação (art. 134 num. 3 e 4).

A Suíça, por sua vez, possui uma limitação aplicável à mineração de textos e dados na pesquisa, mas com suas próprias restrições: o Art 24.d- 1 permite a cópia de obras protegidas para usos que sejam incidentais a processos técnicos, permitindo, ainda, a retenção do material para fins de arquivamento. Contudo, o artigo possui algumas restrições relevantes: a reprodução não é autorizada, apenas obras acessadas de forma lícita podem ser permitidas e, finalmente, programas de computador estão excluídos desta limitação (Suíça, 2022).¹⁵¹ Tratam-se de restrições mais severas, que garantiram ao país a categoria laranja por Flynn et al (2022).

Nota-se que, em diversos desses casos, o que importa – ressalvados os demais requisitos legais – é determinar quais direitos exclusivos estão envolvidos e qual o propósito por trás do uso, sem que se especifique o meio técnico para tanto. Nesse caso, pode-se interpretar os

it shall be accompanied by an acknowledgement of the title of the work and its authorship except where the work is incidentally included in a broadcast; (...).” (Nigéria, 2002).

¹⁵⁰ “**29 Research and private study.**

(1) Fair dealing with a work for the purposes of research for a non-commercial purpose does not infringe any copyright in the work provided that it is accompanied by a sufficient acknowledgement.]

(1B) No acknowledgement is required in connection with fair dealing for the purposes mentioned in subsection (1) where this would be impossible for reasons of practicality or otherwise.

(1C) Fair dealing with a work for the purposes of private study does not infringe any copyright in the work.]

(...)

(3) Copying by a person other than the researcher or student himself is not fair dealing if—

(a) in the case of a librarian, or a person acting on behalf of a librarian, that

person does anything which is not permitted under section [copying by librarians: single copies of published works), or

(b) in any other case, the person doing the copying knows or has reason to believe that it will result in copies of substantially the same material being provided to more than one person at substantially the same time and for substantially the same purpose.” (Reino Unido, 2023).

¹⁵¹ “Art 24.d- 1. For the purposes of scientific research, it is permissible to reproduce a work if the copying is due to the use of a technical process and if the works to be copied can be lawfully accessed. 2. On conclusion of the scientific research, the copies made in accordance with this article may be retained for archiving and backup purposes. 3. This article does not apply to the copying of computer programs” (Suíça, 2022).

processos de mineração de textos e dados enquanto meio para um fim previsto em lei, sendo cobertos por essas provisões como qualquer outra técnica de pesquisa.

3.2.2. Japão

O Japão é mencionado de forma individualizada neste contexto por algumas razões. A primeira é o fato de o país possuir a mais antiga limitação de Direito Autoral que menciona a mineração de textos e dados de forma explícita em lei, precedendo todos os demais exemplos legislativos elencados neste trabalho. A segunda reside na constatação de que tal limitação passou por diversas transformações ao longo dos anos, evoluindo para uma redação singular em relação a outros países que possuem limitações de Direito Autoral nesse sentido – figurando, até hoje, como o único exemplo de limitação para mineração de textos e dados que aplica, de forma mais expressa, a noção de uso não-expressivo da mineração de dados, conforme destacada por SAG (2019) e Carroll (2022).

Mais especificamente, o *Copyright Act* japonês introduziu, por meio do *Act no. 73* de 2009,¹⁵² uma limitação à proteção de bancos de dados por Direito Autoral específica para fins como a “análise de dados” – termo que inclui, como exemplo, atos de extração, comparação, classificação e outros relacionados. Isto implicava em uma permissão não apenas para a mineração de textos e dados em si, mas também para todos os atos necessários para realizá-la.

No que se refere aos direitos envolvidos, apenas a reprodução e a adaptação do material protegido eram autorizadas, sem menção à possibilidade de uso na divulgação de resultados de pesquisa, por exemplo. Observa-se, também, que a limitação excluía a possibilidade de uso de bancos de dados que já são destinados à análise de dados. Ainda, o trecho “na medida considerada necessária” aparecia como uma restrição mais aberta, impondo uma proibição à cópia excessiva sem comprometer a acessibilidade do material necessário para análise. O que chama mais atenção, porém, é que o texto não fazia distinção entre o uso comercial ou não-comercial da análise de dados: ambas as possibilidades, em tese, eram permitidas.

¹⁵² (Reprodução, etc. para análise de informação)

Artigo 47-7. Na medida considerada necessária, é permitido gravar uma obra em uma mídia de gravação ou fazer uma adaptação de uma obra (incluindo a gravação de um trabalho derivado criado pela adaptação) se o objetivo de fazê-lo for a análise de dados (que significa a extração, comparação, classificação ou outra análise estatística de dados de idioma, som ou imagem, ou outros elementos dos quais um grande número de obras ou um grande volume de dados é composto; o mesmo se aplica a seguir neste artigo) por meio de um computador; desde que, no entanto, isso não se aplique a trabalhos de banco de dados compilados para uso por pessoas que realizam análises de dados. (Japão, 2009. Tradução feita pelo autor). Para acessar a fonte oficial, em inglês, ver a atualização de 14 de julho de 2015: <https://www.japaneselawtranslation.go.jp/en/laws/view/2506>. Acesso em 21 nov. 2023.

Este texto legal ainda sobreviveria por quase uma década, durante a qual o Japão manifestou um crescente interesse em incluir atividades intensivas em dados em seus pacotes de políticas públicas de inovação. No documento *Information Economy*, de autoria do Ministério da Economia, Comércio e Indústria do Japão, o governo japonês estabelecia regras e diretrizes com o intuito de “facilitar a utilização de dados e considerar respostas às novas tecnologias emergentes de *blockchain* e ao conceito de economia do compartilhamento, a fim de desenvolver um melhor ambiente de inovação para responder às mudanças no ambiente competitivo das empresas e às mudanças nas estruturas industriais lideradas pelos avanços na IoT e IA” (Japão, 2024). Isto incluiu a criação do IoT Acceleration Consortium, com um grupo de trabalho dedicado a analisar questões relacionadas à promoção de distribuição de dados, destacando questões relativas a transações de dados entre empresas (Japão, 2016, p. 1).

Já a Estratégia de Crescimento do Japão de 2017 enumera diversas iniciativas adotadas ou em fase de implementação (implementadas ou em progresso) visando estimular avanços em inovação, incluindo a remoção de gargalos gerados por incerteza regulatória. Mais especificamente, destacamos a iniciativa “Society 5.0”, focada em estimular a inovação em campos de fronteira tecnológica – o que incluiria inteligência artificial, Internet das Coisas, robôs e *Big Data* em diversos setores – e que incluía reformas regulatórias entre suas iniciativas (Japão, 2017c, p. 13).

O *Programa Estratégico de Propriedade Intelectual de 2017*, por sua vez, mostra mais claramente o papel da propriedade intelectual no cenário da Society 5.0, destacando O programa destaca não apenas a necessidade de estimular o uso da PI por pequenas empresas e reforçar redes de colaboração setor privado-academia (Japão, 2017a, p. 56-57), como também aborda diretamente o Direito Autoral: o documento manifesta preocupação com a impossibilidade de obter licenças de uso de todos os titulares de direitos, manifestando preferência por um sistema de “limitações flexíveis de direitos”, visando equilibrar clareza jurídica e flexibilidade. Isto inclui a regulação de três tipos de atividade: (1) atividades fora do escopo do uso normal de obras protegidas e que geralmente não prejudicam os interesses dos titulares de direitos; (2) atividades fora do escopo do uso normal de obras protegidas e que geralmente causam apenas uma pequena desvantagem aos titulares de direitos, e (3) atividades pelas quais se espera promover o uso de obras protegidas por direitos autorais para políticas com fins de interesse público (*Ibid.*, p. 8-9).

Frise-se que tais ações não ocorreram em um vácuo, desenvolvendo-se de forma não apenas simultânea às conversas realizadas no Parlamento Europeu (a serem tratadas na próxima

seção), mas em comunicação com a região: pelo que se observa do comunicado do Ministério da Economia japonês, intitulado *Japan-EU Joint Press Statement on Facilitating the Free Flow of Data Issued* (Japão, 2017b), o Japão e a União Europeia firmaram uma parceria no sentido de promover um maior fluxo de dados entre as regiões. Isto incluiria a criação de espaços de intercâmbio de experiências e a promoção de um ambiente institucional que garantisse maior proteção dos dados e, ao mesmo tempo, uma circulação mais livre dos mesmos. A declaração também ressalta a importância em envolver pequenas e médias empresas (PME) na cooperação industrial entre o Japão e a EU. O documento se encerra anunciando a assinatura de um memorando de entendimento entre o IoT Acceleration Consortium e a European Alliance for Internet of Things Innovation (AIOTI).

Dentre as soluções para este problema de flexibilidade, chegou-se a explorar a possibilidade de introduzir uma cláusula geral nos moldes norte-americanos (Japão, 2008). Entretanto, o governo japonês concluiu que esta opção poderia trazer insegurança excessiva aos titulares, optando por uma solução alternativa: criar cláusulas fechadas, porém bastante flexíveis (Japão, 2017c). A culminação deste processo se daria no *Act* nº 30 de 2018, aplicável ao *Copyright Act* japonês (Japão, 2019) – o qual trouxe inovações às limitações e exceções para mineração de textos e dados que, até o momento, ainda não encontram paralelo em outras legislações. A principal delas foi o art. 30-4, que previa:

(Exploração sem o propósito de desfrutar dos pensamentos ou sentimentos expressos em uma obra)

Artigo 30-4 É permitido explorar uma obra, de qualquer forma e na medida considerada necessária, em qualquer um dos casos a seguir, **ou em qualquer outro caso em que não seja o propósito de uma pessoa desfrutar pessoalmente ou fazer com que outra pessoa desfrute dos pensamentos ou sentimentos expressos nessa obra**; desde que, no entanto, isso não se aplique se a ação prejudicar injustificadamente os interesses do titular dos direitos autorais diante da natureza ou do propósito da obra ou das circunstâncias de sua exploração:

(i) se for feito para uso em testes para desenvolver ou colocar em uso prática tecnologia relacionada à gravação de sons ou imagens de uma obra ou outra exploração semelhante;

(ii) se for feito para uso em análise de dados (ou seja, a extração, comparação, classificação ou outra análise estatística do idioma, sons, imagens ou outros dados elementares constituintes de um grande número de obras ou de um grande volume de outros dados; o mesmo se aplica ao Artigo 47-5, parágrafo (1), item (ii));

(iii) se for explorada no decurso do processamento de dados computacionais ou explorada de maneira que não envolva o que está expresso na obra sendo percebido pelos sentidos humanos (para obras de programação de computador, tal exploração exclui a execução da obra em um computador), além do estabelecido nos dois itens anteriores. (Japão, 2019. Tradução e destaques feitos pelo autor).¹⁵³¹⁵⁴

¹⁵³ A tradução oficial em inglês se encontra disponível em: https://www.japaneselawtranslation.go.jp/en/laws/view/3379#je_ch2sc3sb5at4. Acesso em 21 nov. 2023.

¹⁵⁴ Foram feitas novas emendas na lei de Direito Autoral japonesa desde então. O site da OMPI conta com a tradução oficial mais recente (inserir link). Nota-se que o Art. 30-4 recebeu alterações de cunho meramente

Aqui, destaca-se o principal diferencial deste dispositivo em relação aos demais modelos jurídicos observados em outros países: a incorporação, em lei, da mineração de textos e dados como uso não-expressivo, que tem como propósito extrair e utilizar apenas dados “brutos”, ou, de outra forma, elementos que não constituem expressão original (European Copyright Society, 2017; Sag, 2019; Carroll, 2022). Contudo, observa-se que este artigo parece não se encaixar adequadamente à hipótese em que um projeto necessita mostrar partes da obra como parte do seu objetivo, ainda que este não seja o objetivo principal do uso e “na medida considerada necessária” (a exemplo do caso Google Books, visto anteriormente).

Destaca-se, ainda, que esta não é a única limitação aplicável à mineração de textos e dados nesta emenda: o Artigo 47-4 permite a exploração de qualquer obra disponibilizada para uso em um computador, desde que o objetivo seja disponibilizar esse trabalho para ser explorado incidentalmente e de maneira mais eficiente em um dado computador – a exemplo de reproduções de obras realizadas quando do processamento de dados contidos nesses trabalhos.¹⁵⁵

redacional, sem alterar o conteúdo por si só – razão pela qual mantivemos a versão de 2018 como objeto de análise neste trabalho.

¹⁵⁵ (Exploitation of Works Incidental to the Exploitation of Works on a Computer)

Article 47-4 (1) A person may exploit a work that is made available to be exploited on a computer (this is inclusive of exploitation using information and communication technologies; the same applies hereinafter in this Article), in any way and to the extent considered to be necessary, in one of the following cases or in any similar case in which the purpose is to make that work available to be exploited incidentally on a computer so that the relevant work can be smoothly or efficiently exploited on that computer; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:

(i) if the person is exploiting a work on a computer using a copy of that work or is exploiting a work transmitted as a wireless communications or wired telecommunications after having received such a transmission, and if, in the course of the data processing that the computer does in order for that work to be exploited, the person records the work on the recording medium of that computer so that it can smoothly and efficiently perform that data processing;

(ii) if a person that, in the course of trade, makes available an automatic public transmission server for another person to use for automatic public transmissions records a work that has been made available for automatic public transmission on a recording medium in order to prevent delays or failures of that other person's automatic public transmissions or in order to efficiently transmit a work that has been made available for transmissions so as to relay automatic public transmissions of that work;

(iii) if the person is providing data by a means that applies information or communication technologies, and records a work on a recording medium or adapts it in order to undertake the computerized data processing that is necessary to prepare to provide that data smoothly and efficiently.

(2) A person may exploit a work that is made available to be exploited on a computer, in any way and to the extent considered to be necessary, in one of the following cases or in any similar case in which the purpose is to maintain or recover the possibility of exploiting a work on that computer; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:

(i) if, in order to perform maintenance or repairs on a device that has a recording medium built into it, the person temporarily records a work that has been recorded on the recording medium that has been built in to that device (hereinafter in this item and the following item referred to as a "built-in recording medium") on a recording medium other than the built-in recording medium, and then re-records that work onto the built-in recording medium after the maintenance or repairs;

Finalmente, destaca-se o Artigo 47-5, que permite a exploração de obras apresentadas ao público em certos casos que contribuam para a criação de conhecimento ou de informação através do processamento de dados – desde que esta exploração seja dentro dos limites considerados necessários, seja de caráter proporcionalmente mínimo em relação ao total da obra utilizada e, como de praxe, não viole, de forma consciente, os interesses legítimos dos titulares de direitos de autor face à natureza do trabalho e da exploração. Esses casos envolvem (1) a realização de análise de dados via computador e a divulgação dos resultados da mesma; (2) atos previstos pelo governo que contribuam para aprimorar a qualidade de vida dos japoneses através da criação desse novo conhecimento ou informação. Ao mesmo tempo, cobre-se a omissão do Art. 30-4 em relação à divulgação de conteúdo protegido: o Artigo 47-5 determina que qualquer pessoa responsável pela coleta, organização e fornecimento de informação para a preparação dos usos previstos anteriormente tem permissão expressa para reproduzir ou transmitir as obras envolvidas, dentro dos limites estabelecidos na lei.¹⁵⁶

(ii) if, in order to replace a device that has a recording medium built into it with another device with the same functions, the person temporarily records a work that has been recorded onto that device's built-in recording medium onto a recording medium other than that built-in recording medium, and then records that work on the built-in recording medium of the device with same functions;

(iii) if a person that, in the course of trade, makes available an automatic public transmission server for another person to use for automatic public transmissions records a work on a recording medium for use in recovery in the event that the copy of the work that has been made available for automatic public transmission by that automatic public transmission server is lost or damaged (Japão, 2018).

¹⁵⁶ (Minor Exploitation Incidental to Computerized Data Processing and the Provision of the Results Thereof)

Article 47-5 (1) A person undertaking an action as set forth in one of the following items that contributes to facilitating the exploitation of a work by creating new knowledge or information through computerized data processing (this includes a person undertaking a part of such an action; limited to one doing so in accordance with the standards prescribed by Cabinet Order) may exploit a work that has been made available or presented to the public (this includes a work that has been made available for transmission; the same applies hereinafter in this Article) (hereinafter in this Article and Article 47-6, paragraph (2), item (ii) referred to as "available or presented work") (limited to a publicized work or a work made available for transmission), in any way and to the extent considered to be necessary in light of the purpose of the action set forth in the relevant item, when exploiting it incidental to the undertaking of that action (limited to exploitation that is minor in light of the percentage it constitutes of the part of the available or presented work that has been provided for exploitation, the volume of the part of that work that has been provided for exploitation, the accuracy of indications made at the time it was provided for exploitation, and other elements; hereinafter in this Article referred to as "minor exploitation"); provided, however, that this does not apply if the person undertakes that minor exploitation knowing that the available or presented work's having been made available or presented to the public constitutes copyright infringement (for a work made available or presented to the public abroad, this means that the action would constitute copyright infringement if it took place in Japan), nor does it apply if the action would otherwise unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the available or presented work or the circumstances of its minor exploitation:

(i) using a computer to search for the title or author name of a work in which information that a person is searching for (hereinafter in this item referred to as "information being searched for") has been recorded, for the transmitter identification code (meaning the letters, numbers, symbols, or any other code by which the transmitter of an automatic public transmission is identified) associated with information being searched for that has been made available for transmission, or for any other information concerning the identification or location of information being searched for; and making the results of that search available;

(ii) undertaking computerized data analysis and furnishing the results of that analysis;

Para além do texto das limitações em si, há ainda uma ressalva em relação a dispositivos anti-burla no que tange à satisfação dessas limitações. Mais especificamente, o Artigo 113, parágrafos (6) da lei japonesa, ao proibir que se contornem medidas de proteção tecnológica, cria uma exceção na qual tais procedimentos são possíveis para fins de pesquisa, desde que preservados os interesses do titular.¹⁵⁷

Mais recentemente, em meio ao fortalecimento da IA Generativa, o governo japonês reafirmou sua postura de abertura à mineração de textos e dados. No documento “AIと著作権” (IA e direitos autorais, em japonês), a Agência de Assuntos Culturais do Japão reforçou a ideia de que ideias devem ser mantidas fora do escopo da proteção por direitos autorais a fim de permitir o florescimento de atividades criativas, e que o “estilo artístico” se enquadraria no campo das ideias não-protegidas (Japão, 2020, p. 10).¹⁵⁸ Por outro lado, o documento considera que o uso de obras protegidas em IA Generativa deve ser dividido em duas etapas: o treinamento de uma inteligência artificial, onde se considera que o uso de material protegido é inerentemente não-expressivo, e a fase de geração, onde pode haver uso expressivo (portanto, fora do escopo da limitação). Contudo, o parâmetro para avaliação permaneceria sendo a comparação do produto da IA com a obra original que se alega ter sido utilizada. Se não houver similaridade suficiente, ou se apenas elementos factuais ou não-originais estiverem presentes, a infração não se

(iii) an action that Cabinet Order prescribes as contributing to increased convenience in the lives of the citizenry by creating new knowledge or information through computerized data processing and making the results of this available, beyond what is set forth in the preceding two items.

(2) A person that prepares to undertake an action set forth in one of the items of the preceding paragraph (limited to a person that collects, organizes, and provides information in preparation to undertake the action in accordance with the standards prescribed by Cabinet Order) may reproduce or make public transmissions of an available or presented work (or make the relevant work available for transmission, if such transmission is being made via an automatic public transmission; the same applies in this paragraph and Article 47-6, paragraph (2), item (ii)) or distribute copies thereof, to the extent considered to be necessary in order to prepare for minor exploitation under the preceding paragraph; provided, however, that this does not apply if the action would otherwise unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the available or presented work, the number of copies that would be reproduced or distributed, or the circumstances of the reproduction, public transmission, or distribution (Japão, 2018).

¹⁵⁷ “Article 113 (...)

(6) The circumvention of technological exploitation restriction measures (meaning making it possible to view or listen to a work, etc. that technological exploitation restriction measures have been used to restrict from being view or listened to, by hindering the effect of those technological exploitation restriction measures (unless this is done based on the intent of the owner of the Copyright; Related Matters.); the same applies in the following paragraph and Article 120-2, items (i) and (ii)) is deemed to constitute infringement of the copyright, print rights, or neighboring rights connected to those technological exploitation restriction measures unless this is done within a scope that is justified for the purpose of research or technological development in connection with technological exploitation restriction measures, or otherwise does not harm the interests of the owner of the Copyright; Related Matters.” (Japão, 2018).

¹⁵⁸ “A proteção de “ideias” abstratas pode impedir novas atividades criativas e expressivas posteriores (...). É mais provável que as “ideias” abstratas possam ser usadas livremente, o que leva a uma maior variedade e abundância de criação e expressão concretas. (...) O estilo do artista em si não é protegido (mesmo que o estilo seja semelhante, isso não constitui violação de direitos autorais)” (Japão, 2020, p. 10. Tradução feita pelo autor).

observará (*Ibid.*, p. 34; 43-45).¹⁵⁹ A exceção para esta regra seria a existência de um mercado estabelecido para licenciamento de uma determinada obra em análise de dados, em cujo caso a limitação prevista na lei japonesa não se aplicaria.¹⁶⁰ Adota-se, portanto, um modelo conceitual de separação em etapas do processo de mineração de textos e dados, no qual um tipo de uso (o *input*) sempre será lícito, enquanto o *output* dependerá do caso concreto.

Tais regulações, quando observadas em conjunto, sinalizam, por um lado, um aparente paradoxo: o Japão, ao se afastar do modelo mais aberto visto nos EUA e optar por uma lista fechada, mas flexível, parece criar um tipo peculiar de L&E – um que dita explicitamente a regra dos três passos ao mesmo tempo em que testa seus limites do que se considera um “certo caso especial”, com hipóteses legais que comportam uma potencial multiplicidade de cenários.

Dadas as considerações sobre o papel das limitações e exceções feitas ao final do capítulo anterior, seria possível interpretar este movimento regulatório como um posicionamento no sentido de entender as L&E para além de uma ótica excepcionalista, usando-as não como freios, mas como reconhecimento dos limites naturais da propriedade intelectual. A análise de Ueno (2021, p. 150-151) parece corroborar esta questão: para o autor, o art. 30-4 representa a consolidação de uma noção pela qual direitos de autor e conexos só têm alcance sobre o valor no uso de uma obra “enquanto obra” – tal como o direito marcário se destina a proteger apenas os sinais, siglas e símbolos que são utilizados “enquanto marcas”, em comércio. Em outras palavras, o Direito Autoral se destinaria a proteger o valor pessoal e econômico estritamente decorrente da apreciação estética de uma obra artístico-literária: usos que escapam desta finalidade, como a análise de dados factuais contidos nessas obras, não encontram justificativa neste sistema.¹⁶¹

¹⁵⁹ “É necessário considerar separadamente o estágio no qual a obra é inserida e o estágio no qual ela é produzida (...).O estágio de entrada é considerado um uso que não aproveita a expressão da obra, e que normalmente não se considera que haja prejuízo para o detentor dos direitos autorais”); 43-45 (“O fato de o uso de IA para gerar imagens etc. constituir ou não violação de direitos autorais é determinado da mesma forma que em casos convencionais, como quando uma pessoa faz um desenho sem usar IA (...) Se o produto de IA não tiver “similaridade com” ou “se basear em” um trabalho existente, ele não constituirá uma violação dos direitos autorais do trabalho existente e poderá ser usado sem a permissão do detentor dos direitos autorais, nos termos da Lei de Direitos Autorais”) (Japão, 2020, p. 34).

¹⁶⁰ “Se as obras para as quais foi estabelecido um mercado de licenças para análise de informações forem disponibilizadas para análise de informações sem autorização nos termos das disposições de limitação de direitos, os interesses dos detentores de direitos autorais poderão ser indevidamente prejudicados.

(...)

O fato de uma obra se enquadrar ou não na ressalva será, em última análise, julgado em um contexto judicial, caso a caso, em termos de conflito com o mercado para o uso da obra do detentor dos direitos autorais ou de impedimento dos possíveis canais de vendas futuras da obra” (Japão, 2020, p. 39-40).

¹⁶¹ “A copyrighted work is one that satisfies an intellectual or emotional demand through the enjoyment of it (e.g. listening to music, looking at a picture, watching a movie, reading a novel or executing a computer program), and people pay compensation for the enjoyment of a work. The reason why copyright law grants copyright to an author

Por outro lado, a menção contínua da propriedade intelectual associada à estratégia de inovação, bem como o emprego de múltiplos usos livres associados a atividades de cunho técnico ou de imediata reversão para o interesse social, pode levar a um entendimento complementar: a de que estas regulações representariam um compromisso em relação a um programa maior de políticas públicas voltadas para o estímulo ao desenvolvimento de tecnologia de ponta. Neste cenário, o Direito Autoral não aparece como um objeto insular ou como um regime que se justifica por si só, mas como mais um componente institucional dentro do sistema japonês de inovação, pautado na garantia de segurança jurídica e flexibilidade para usos diretamente ligados a possibilidades de crescimento econômico.

No geral, o Japão mantém um conjunto de limitações e exceções bastante permissivo para a mineração de textos e dados para quaisquer fins no que tange ao escopo dos direitos exclusivos (todos) e do propósito (qualquer um que não resulte no uso direto da expressão artística no resultado final do processo). Isto garante ao país o grau mais alto de abertura na classificação de Flynn et al (2021), o que representa, a primeira vista, amplas oportunidades para pesquisadores sem ignorar completamente possíveis problemas com a reprodução e comercialização indevida de obras – embora mantenha uma postura cautelosa quanto a definir a posição de usos derivados de IA neste cenário. Além disso, cabe observar que não há, à primeira vista, qualquer previsão legal sobre a possibilidade de abdicar dessas limitações pela via contratual, o que dá contornos ainda nebulosos ao escopo total de aplicabilidade desta limitação.

3.2.3. União Europeia

A evolução das limitações para mineração de textos e dados na Europa possui suas próprias peculiaridades. Aqui, destaca-se um breve (e recente) histórico das L&E no campo da prospecção de dados para pesquisa dentro da Europa, o que inclui tanto iniciativas em países isolados como iniciativas dentro da União Europeia: das limitações para pesquisa nas Diretivas 96/9/CE e 2001/29/CE, até a criação de uma nova e mais abrangente limitação no âmbito do Mercado Único Digital, como veremos adiante.

Nesse sentido, embora o Japão tenha sido o primeiro a mencionar a mineração de textos e dados de forma direta, a União Europeia já havia determinado a criação de uma limitação

is to guarantee the opportunity to receive compensation directly or indirectly from those who want to enjoy the work (...). In light of the above-mentioned theory (...), it seems that the nature of Art. 30-4 can be substantially considered an ‘internal’ rather than ‘external’ limit to copyright as it strikes a balance between copyright and other fundamental rights such as freedom of expression. Therefore, Art. 30-4 can be substantially regarded as an instrument to determine the inherent scope of copyright, which might be similar to the theory of ‘a use as a (trade)mark’ or the functions of trademark, even though Art. 30-4 is formally provided for, not as the provision on the scope of copyright, but one of the provisions on copyright exceptions.” (Ueno, 2021, p. 150-151).

aplicável à coleta de dados para fins de pesquisa no âmbito do direito *sui generis* sobre o conteúdo de bases de dados, regulado na Diretiva 96/9/CE. Mais especificamente, o art. 9 da Diretiva permite que usuários legítimos de uma base colocada à disposição do público extraiam e reutilizem uma parte substancial desse repositório para fins de pesquisa, desde que haja indicação da fonte, tal uso se dê na medida considerada necessária e o uso tenha finalidade não-comercial.¹⁶² Derclaye (2008, p. 48) observa, entretanto, que tal disposição é opcional: os países-membros não foram obrigados a implementá-la.

Ainda no âmbito desta diretiva, contudo, a autora (*Ibid.*, p. 92-94) também nota que a Corte de Justiça da União Europeia, em diversos casos posteriores à promulgação da Diretiva, procedeu a uma forma peculiar de expansão das limitações ao direito *sui generis*: não impondo casos especiais, mas reduzindo o escopo do próprio objeto de proteção desse instituto – notadamente, ao estipular, por exemplo, que a proteção pelo instituto não se aplica se o investimento realizado na coleta, verificação ou apresentação dos dados for simultâneo e inseparável da criação desses dados, a fim de evitar a monopolização de conhecimento. Exemplos de conjuntos dentro desta definição incluiriam grades de programação e listas de clientes (e as respectivas informações).

Anos mais tarde, a Diretiva 2001/29/CE (a Diretiva InfoSoc), ao tratar da harmonização das legislações europeias de Direito Autoral na sociedade da informação, abordou, em seu preâmbulo, a possibilidade de seus Estados-Membros implementarem limitações aplicáveis não apenas a cópias transitórias de obras como, mais uma vez, de limitações destinadas à pesquisa científica.¹⁶³ Isto se materializou no artigo 5, (1) e (3), onde se permite a reprodução transitória

¹⁶² “Artigo 9º

Excepções ao direito *sui generis*

Os Estados-membros podem prever que o utilizador legítimo de uma base de dados posta à disposição do público, seja por que meio for, possa, sem autorização do fabricante da base extrair e/ou reutilizar uma parte substancial do seu conteúdo:

(...)

b) Sempre que se trate de uma extracção para fins de ilustração didáctica ou de investigação científica, desde que indique a fonte e na medida em que tal se justifique pelo objectivo não comercial a atingir (...)” (União Europeia, 1996).

¹⁶³ “(33) O direito exclusivo de reprodução deve ser sujeito a uma excepção para permitir certos actos de reprodução temporária, que são reproduções transitórias ou pontuais, constituindo parte integrante e essencial de um processo tecnológico efectuado com o único objectivo de possibilitar, quer uma transmissão eficaz numa rede entre terceiros por parte de um intermediário, quer a utilização legítima de uma obra ou de outros materiais protegidos. Os actos de reprodução em questão não deverão ter, em si, qualquer valor económico. Desde que satisfeitas essas condições, tal excepção abrange igualmente os actos que possibilitam a navegação (“browsing”) e os actos de armazenagem temporária (“caching”), incluindo os que permitem o funcionamento eficaz dos sistemas de transmissão, desde que o intermediário não altere o conteúdo da transmissão e não interfira com o legítimo emprego da tecnologia, tal como generalizadamente reconhecido e praticado pela indústria, para obter dados sobre a utilização da informação. Uma utilização deve ser considerada legítima se tiver sido autorizada pelo titular de direitos e não estiver limitada por lei.

para fins não-econômicos, bem como se repete, em larga medida, a limitação já existente na diretiva de bases de dados, com o bônus de desobrigar usuários de citar a fonte quando isto não é possível.¹⁶⁴ Nota-se, também, que o preâmbulo do documento deixa claro que o critério para definição de pesquisa não-comercial é tão somente a finalidade, e não o caráter comercial do usuário.¹⁶⁵

Contudo, a menção explícita a técnicas de mineração de textos e dados dentro da Europa começaria a se fazer mais presente a partir da década seguinte, no âmbito de países específicos. Um deles é a França, que, em 2016, permitia explicitamente o uso dessas técnicas, embora com restrições significativas: aplicava-se apenas a materiais contidos em publicações científicas já divulgadas e o uso comercial era expressamente proibido.¹⁶⁶ A Alemanha constitui outro exemplo: em 2017, o país tinha uma limitação específica para a coleta e análise de dados na pesquisa científica, esta bastante permissiva: permitia a reprodução de material protegido para normalização, estruturação e categorização de um *corpus* para análise, bem como a sua reprodução limitada para colaboradores e para outros pesquisadores, para fins de verificação da qualidade da pesquisa. Aqui, o uso comercial também não era permitido, e havia um adendo:

(34) Deve ser dada aos Estados-Membros a opção de preverem certas exceções e limitações em determinados casos, nomeadamente para fins de ensino ou de investigação científica, a favor de instituições públicas como bibliotecas e arquivos, para efeitos de notícias, citações, para utilização por pessoas deficientes, para utilização relacionada com a segurança pública e para utilização em processos administrativos e judiciais.” (União Europeia, 2001).

¹⁶⁴ Artigo 5.o

Exceções e limitações

1. Os actos de reprodução temporária referidos no artigo 2.o, que sejam transitórios ou episódicos, que constituam parte integrante e essencial de um processo tecnológico e cujo único objectivo seja permitir:

- a) Uma transmissão numa rede entre terceiros por parte de um intermediário, ou
- b) Uma utilização legítima de uma obra ou de outro material a realizar, e que não tenham, em si, significado económico, estão excluídos do direito de reprodução previsto no artigo 2.o.

(...)

3. Os Estados-Membros podem prever exceções ou limitações aos direitos previstos nos artigos 2.o e 3.o nos seguintes casos:

- a) Utilização unicamente com fins de ilustração para efeitos de ensino ou investigação científica, desde que seja indicada, excepto quando tal se revele impossível, a fonte, incluindo o nome do autor e, na medida justificada pelo objectivo não comercial que se pretende atingir (...).” (União Europeia, 2001).

¹⁶⁵ “(42) Na aplicação da excepção ou limitação para efeitos de investigação pedagógica e científica não comercial, incluindo o ensino à distância, o carácter não comercial da actividade em questão deverá ser determinado por essa actividade propriamente dita. A estrutura organizativa e os meios de financiamento do estabelecimento em causa não são factores decisivos a esse respeito.” (União Europeia, 2001).

¹⁶⁶ “Article L122-5

Lorsque l'oeuvre a été divulguée, l'auteur ne peut interdire :

(...)

10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche” (França, 2016).

finda a pesquisa, havia a obrigação de destruir o material e removê-lo do acesso ao público, embora o corpus pudesse ser copiado e armazenado por uma entidade de conservação.¹⁶⁷

Entretanto, nos últimos anos, a União Europeia tem manifestado uma preocupação maior com a harmonização das regulações de seus países-membros no que diz respeito ao controle do fluxo de dados e à facilitação do desenvolvimento e usufruto de bens e serviços baseados em tecnologias intensivas em dados, bem como o setor de internet. Estes esforços foram combinados na iniciativa do Mercado Único Digital, ou Digital Single Market.¹⁶⁸

Dois documentos recebem destaque aqui. O primeiro, a Comunicação ao Parlamento Europeu *Construir uma economia europeia dos dados*, nas seções 2 e 3.2 (“Livre circulação de dados” e “Acesso limitado aos dados”, respectivamente), destaca a circulação livre, segura e confiável de dados como sendo crucial para garantir o livre trânsito de mercadorias, trabalhadores, serviços e capital dentro do bloco. Mais especificamente, entende-se pela necessidade não somente de capacitar empresas para melhor valorizar e proteger seus dados, mas também de incentivar a geração de empregos e o desenvolvimento econômico através do estímulo ao compartilhamento de dados e da remoção de barreiras – dentre as quais o fato de muitos geradores de dados os manterem isolados e guardados em sigilo, comprometendo o acesso e transparência dos mesmos (União Europeia, 2017).

Além disto, na seção 3.4 (“Situação de facto”) o documento também aborda as relações entre detentores de dados e usuários, bem como negócios de menor porte, ao afirmar que os

¹⁶⁷ Section 60d Text and data mining

(1) In order to enable the automatic analysis of large numbers of works (source material) for scientific research, it shall be permissible

1. to reproduce the source material, including automatically and systematically, in order to create, particularly by means of normalisation, structuring and categorisation, a corpus which can be analysed and

2. to make the corpus available to the public for a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research.

In such cases, the user may only pursue non-commercial purposes. (2) If database works are used pursuant to subsection (1), this shall constitute customary use in accordance with section 55a, first sentence. If insubstantial parts of databases are used pursuant to subsection (1), this shall be deemed consistent with the normal utilisation of the database and with the legitimate interests of the producer of the database within the meaning of section 87b (1), second sentence, and section 87e. (3) Once the research work has been completed, the corpus and the reproductions of the source material shall be deleted; they may no longer be made available to the public. It shall, however, be permissible to transmit the corpus and the reproductions of the source material to the institutions referred to in sections 60e and 60f for the purpose of long-term storage. (Alemanha, 2017).

¹⁶⁸ Segundo a Comissão Europeia, o Mercado Único Digital surge como forma de harmonizar a regulação sobre dados e Internet no âmbito da União Europeia, removendo obstáculos ao total usufruto de bens e serviços disponíveis digitalmente (por usuários, governos e empresas), e retirando barreiras regulamentares para o espaço de atuação das empresas do setor da Internet e das *start-ups*. A proposta se baseia em três pilares: maior acesso dos consumidores e empresas aos bens disponíveis *on-line*; um ambiente mais acolhedor ao crescimento de redes e serviços digitais e a economia digital como componente vital ao crescimento econômico (União Europeia, 2019b). Para mais informações, ver: https://ec.europa.eu/commission/priorities/digital-single-market_pt. Acesso em 29 out. 2024.

fabricantes ou prestadores de serviços se tornam proprietários “de fato” dos dados produzidos por máquinas: nesse contexto, afirma-se que nem mesmo o próprio usuário se tornaria incapaz de autorizar o uso, por terceiros, de conteúdo gerado a partir de programas ou dispositivos adquiridos e usados por este usuário. Além disso, o documento prossegue apontando que as lacunas regulatórias presentes à época permitiriam a diversos titulares impor acordos exclusivos de exploração dos dados, por meio de contratos abusivos ou meios técnicos (União Europeia, 2017).

Já outra Comunicação ao Parlamento Europeu, intitulada *Rumo a um espaço económico europeu de dados*, destaca a inovação intensiva em dados como componente crucial de tecnologias como inteligência artificial e Internet das Coisas, e um fator-chave para um aumento em ofertas de emprego, inovação e competitividade dentro do bloco. O documento também aponta para a necessidade de democratizar a tecnologia da inteligência artificial para entes comerciais ou não, bem como, no item 3-b (“Acesso à informação científica e sua preservação”), destacou a ciência aberta como essencial para o desenvolvimento científico e social (União Europeia, 2018).

É neste contexto que é aprovada a Diretiva (UE) 2019/790 (União Europeia, 2019a), que visa trazer emendas às Diretivas 96/9/EC e 2001/29/EC – que travam de questões relacionadas à proteção de bases de dados e propriedade intelectual. Em seu preâmbulo, o documento reforça a importância da mineração de textos e dados não somente para a pesquisa científica, como para diversas outras atividades nos setores público e privado, bem como destacou os problemas oriundos usuários da incerteza jurídica sobre o uso de obras protegidas em processos de análise de dados (União Europeia, 2019a). Neste ponto, busca-se expandir as limitações para além do que já era previsto no direito *sui generis* da Diretiva 96/9/CE – o qual já incluía o uso para fins didáticos ou científicos, bem como outros casos específicos.¹⁶⁹

Desta forma, a Diretiva incluiu duas limitações para a mineração de textos e dados a serem implementadas por seus países-membros, aplicáveis não somente ao Direito Autoral,

¹⁶⁹ “Artigo 9º

Excepções ao direito *sui generis*

Os Estados-membros podem prever que o utilizador legítimo de uma base de dados posta à disposição do público, seja por que meio for, possa, sem autorização do fabricante da base extrair e/ou reutilizar uma parte substancial do seu conteúdo:

a) Sempre que se trate de uma extracção para fins particulares do conteúdo de uma base de dados não electrónica;
 b) Sempre que se trate de uma extracção para fins de ilustração didáctica ou de investigação científica, desde que indique a fonte e na medida em que tal se justifique pelo objectivo não comercial a atingir;
 c) Sempre que se trate de uma extracção e/ou de uma reutilização para fins de segurança pública ou para efeitos de um processo administrativo ou judicial” (União Europeia, 1996).

como ao direito *sui generis*. A primeira, o Artigo 3, prevê a possibilidade do uso de material protegido em processos de mineração de textos e dados especificamente no âmbito da pesquisa científica:

Artigo 3.º

Prospecção de textos e dados para fins de investigação científica

1. Os Estados-Membros preveem uma exceção aos direitos previstos no artigo 5.º, alínea a), e no artigo 7.º, n.º 1, da Diretiva 96/9/CE, no artigo 2.º da Diretiva 2001/29/CE, e no artigo 15.º, n.º 1, da presente diretiva no que se refere às reproduções e extrações efetuadas por organismos de investigação e por instituições responsáveis pelo património cultural para a realização de prospecção de textos e dados de obras ou outro material protegido a que tenham acesso legal para efeitos de investigação científica.
2. As cópias de obras ou de outro material protegido efetuadas nos termos do n.º 1 devem ser armazenadas com um nível de segurança adequado e podem ser conservadas para fins de investigação científica, incluindo para a verificação dos resultados da investigação.
3. Os titulares de direitos devem ser autorizados a aplicar medidas para assegurar a segurança e a integridade das redes e bases de dados em que as obras ou outro material protegido são acolhidos. Essas medidas não podem exceder o necessário para alcançar esse objetivo.
4. Os Estados-Membros devem incentivar os titulares de direitos, os organismos de investigação e as instituições responsáveis pelo património cultural a definir melhores práticas previamente acordadas no que se refere à aplicação da obrigação e das medidas a que se referem, respetivamente, os n.ºs 2 e 3 (União Europeia, 2019).

Aqui, algumas considerações merecem destaque. A primeira é que a limitação só se aplica a instituições de pesquisa e de preservação do património cultural, não se estendendo a pesquisadores independentes – o que dá a esta limitação uma posição roxa na classificação de Flynn et al (2022), quando considerada de forma isolada. Ademais, de forma semelhante ao previsto na antiga Diretiva 96/9/CE, esta nova limitação considera apenas usos não-comerciais: embora o artigo em si não torne isto explícito, o Preâmbulo nº 12 da Diretiva enquadra apenas entidades com fins não-lucrativos, excluindo institutos controlados por empresas.¹⁷⁰

Para Ducato e Strowel (2021, p. 12), este nível de restrição quanto a usuários não cobriria toda a extensão de agentes responsáveis pelas atividades de pesquisa: dentre outros, o artigo exclui o trabalho jornalístico, obstaculiza a elaboração de políticas públicas mais sofisticadas¹⁷¹ e, crucialmente, desampara pesquisadores mais novos que porventura se encontrem desvinculados a alguma instituição, dificultando seu engajamento na academia e,

¹⁷⁰ “Apesar de contarem com diferentes formas e estruturas jurídicas, os organismos de investigação dos Estados-Membros têm geralmente em comum o facto de agirem sem fins lucrativos ou no âmbito de uma missão de interesse público reconhecida pelo Estado (...). Pelo contrário, não deverão ser considerados organismos de investigação para efeitos da presente diretiva, os organismos sobre os quais as empresas comerciais têm uma influência decisiva, permitindo às referidas empresas exercer controlo devido a condições estruturais, nomeadamente através da sua qualidade de acionistas ou sócios, o que poderá conduzir a um acesso preferencial aos resultados da investigação” (União Europeia, 2019).

¹⁷¹ Nos termos levantados, por exemplo, por McAfee e Brynjolfsson (2012), como visto no Capítulo 1.

consequentemente, sua exposição – o que, como discutido anteriormente, guarda relação direta com a disponibilidade de oportunidades de emprego neste setor.

Ademais, embora permita a reprodução e o armazenamento de material, inclusive para fins de verificação, não fica claro se o compartilhamento dessas informações é permitido, ainda que justamente para atender a esse propósito de checagem. Também não há permissão para a adaptação desse conteúdo, o que significa que, diferente do Japão e dos casos observados nos Estados Unidos, nem todas as etapas do processo de análise de dados são asseguradas.

Outro ponto de destaque é a opção por permitir apenas a reprodução e extração de obras às quais se teve “acesso lícito” – de forma similar à França, mas destoante de, por exemplo, Alemanha e Japão. A Diretiva define mais especificamente o conceito de “acesso lícito” no item 14 do seu preâmbulo, como aquele referente a conteúdos licenciados, sob políticas de acesso aberto ou livremente disponíveis online.¹⁷² Trata-se de um requisito deveras restritivo, tendo-se em vista a dificuldade de averiguar o status legal de cada um dos materiais coletados dentro de um processo de mineração de dados padrão – o qual, reitera-se, costuma lidar com quantidades de dados muito além da capacidade de processamento humano. Esta dificuldade se agrava ainda mais quando tratamos de mineração para o desenvolvimento de sistemas de IA, uma vez que um modelo treinado permanece inalterado, mesmo que a obra seja removida da base de dados utilizada para o treinamento (Margoni, 2024).¹⁷³

Ademais, esta definição ainda apresenta algumas incertezas. Mais especificamente, para Margoni (*Ibid.*), o dispositivo não torna claro se a licitude exigida no artigo se refere à

¹⁷² “(14) O acesso legal deverá ser entendido como abrangendo o acesso a conteúdos baseados numa política de acesso aberto ou através de acordos contratuais entre titulares de direitos e organismos de investigação ou instituições responsáveis pelo património cultural, tais como assinaturas, ou através de outras vias legais. Por exemplo, no caso de assinaturas feitas por organismos de investigação ou por instituições responsáveis pelo património cultural, considera-se que as pessoas que lhes estão adstritas e se encontram abrangidas por estas assinaturas também deverão ter um acesso legal. O acesso legal deverá abranger igualmente o acesso aos conteúdos livremente disponíveis em linha.” (União Europeia, 2019).

¹⁷³ “The reported line of cases essentially establishes either i) a total ban on unlawful sources (without clarifying how this unlawfulness should be assessed, thus arguably strictly, e.g., ACI Adam), or ii) a presumption of knowledge of the unlawfulness of the source when operating for financial gain purposes (GS Media). Accordingly, research organizations may have to comply with two alternative frameworks. In the first, they are asked to perform a strict assessment of lawfulness for every single piece of content used for TDM, something that may be quite difficult for most content available on the Internet. This framework will almost unavoidably reduce Art. 3 to an ineffective provision (...). By adopting the more lenient case law on linking, it is arguable that research organizations operating for research purposes within Art. 3 do not pursue financial gain and thus are not liable if they have no direct knowledge of the unlawfulness of the source. However, even this second option is not as suitable as it might look. Research organizations’ (presumed) lack of knowledge will cease once they become aware of the unlawfulness of the sources (e.g., via a notification) and they will be under an obligation to remove it (see [GS Media](#) at 53). Here lies the problem with this option. Whereas it is possible to remove a link to an unlawful source ([GS Media](#)), a photograph from a website ([Renckhoff](#)), and arguably also a work from an Art. 3(2) storage dataset, it is simply impossible or (considering statements of theoretical feasibility) extremely costly to remove the information derived from an unlawful source in a trained model.” (Margoni, 2024).

legalidade da fonte utilizada (se as obras obtidas são de origem lícita, como uma biblioteca) ou ao ato de acesso em si (se houve, por exemplo, violação de medidas tecnológicas para a obtenção do material); argumenta-se que ambas as interpretações são possíveis no artigo. Com efeito, Griffiths, Synodinou e Xalabarder (2022) e Margoni (2024) parecem entender este comando como uma medida anti-burla por si própria, a qual visa resguardar medidas tecnológicas de proteção; ao mesmo tempo, ambos os trabalhos se baseiam nos textos de outras Diretivas e de decisões da Corte de Justiça da União Europeia (CJEU) para afirmar que deve-se interpretar o artigo 3 no sentido de considerar como “lícito” o acesso de qualquer conteúdo tornado publicamente disponível. Ainda neste sentido, argumentamos que a exigência de medidas de segurança ou integridade dos repositórios em posse do usuário da limitação serve como medida complementar neste caso, responsabilizando as instituições de pesquisa por eventuais acessos e usos ilícitos por terceiros, de maneira familiar a ressalvas do gênero feitas no *fair use* norte-americano.

De qualquer forma, os termos postos pela exigência de acesso lícito terminam por servir não apenas como uma manutenção da obrigação, por parte das instituições contempladas, de licenciar obras e catálogos, mas também como um instrumento que também permite aos titulares exigirem, como condição para o acesso, uma licença adicional para extração dos dados. Isto, por sua vez, tem o potencial de elevar os custos de acesso e excluir a maioria das instituições de pesquisa, que não possuiriam meios economicamente sustentáveis de manter tais licenças (Griffiths, Synodinou e Xalabarder, 2022, p. 11-12).

Já o Artigo 4 traz uma permissão mais ampla para quaisquer outros casos que extrapolem o Artigo 3. Assim, os países-membros devem permitir o uso de obras protegidas em processos de mineração de textos e dados, pelo tempo que for necessário, para a conclusão do processo – desde que os titulares de direitos não tenham reservado “de maneira adequada” os direitos de fazer, eles mesmos, essa reprodução e extração para TDM:

Artigo 4.º

Exceções ou limitações para a prospeção de textos e dados

1. Os Estados-Membros devem prever uma exceção ou uma limitação aos direitos previstos no artigo 5.º, alínea a), e no artigo 7.º, n.º 1, da Diretiva 96/9/CE, no artigo 2.º da Diretiva 2001/29/CE, no artigo 4.º, n.º 1, alíneas a) e b), da Diretiva 2009/24/CE e no artigo 15.º, n.º 1, da presente diretiva, para as reproduções e as extrações de obras e de outro material protegido legalmente acessíveis para fins de prospeção de textos e dados.
2. As reproduções e extrações efetuadas nos termos do n.º 1 podem ser conservadas enquanto for necessário para fins de prospeção de textos e dados.
3. A exceção ou limitação prevista no n.º 1 é aplicável desde que a utilização de obras e de outro material protegido a que se refere esse número não tenha sido expressamente reservada

pelos respetivos titulares de direitos de forma adequada, em particular por meio de leitura ótica no caso de conteúdos disponibilizados ao público em linha.

4. O presente artigo não prejudica a aplicação do artigo 3.º da presente diretiva.

(União Europeia, 2019)

Aqui, diversas semelhanças com o artigo anterior são notadas: os mesmos direitos de reprodução e extração são mantidos, sem menção à reprodução ou adaptação; a restrição a conteúdos acessados de forma lícita e a ausência da exigência de fins não-comerciais. Porém, duas diferenças fundamentais são observadas.

A primeira delas, em razão da sua aplicabilidade mais geral, é a permissão para quaisquer outros fins e usuários – o que ajuda a complementar, teoricamente, a restrição institucional imposta pelo Artigo 3. Esta “complementação”, contudo, vem com o preço da segunda grande diferença: a possibilidade de ser afastada pelo titular do conteúdo se este se reservar o direito de uso para fins de mineração de textos e dados – seja por contrato, por meios legíveis por máquina ou por qualquer maneira considerada “adequada” (conceito definido no item 18 do preâmbulo da Diretiva),¹⁷⁴ desde que não interfira com o exercício da hipótese do art. 3 da Diretiva ou com o art. 5 da Diretiva InfoSoc.¹⁷⁵

Dentro dessas balizas, portanto, os termos do opt-out são deixados a cargo dos titulares, o que Schirru e Margoni (2023) enxergam como problemático ao observarem a falta de um padrão geral e coerente de aplicação: para além dos critérios estabelecidos por editoras científicas (a serem discutidos no capítulo seguinte), os autores observam diversos termos de uso contraditórios, que não transmitem segurança jurídica suficiente para os licenciandos. Já o caso LAION e.v. vs Kneschke, na Alemanha, surge como produto desta indefinição, levando a questões sobre a validade de uma reserva legal genérica (por exemplo, “todos os direitos reservados”; “para quaisquer fins”, o que foi aceito pela corte responsável pelo caso) e sobre o que exatamente conta como “legível por máquinas” ou não (Keller, 2024b).

Em outra obra, Keller (2024a, p. 9-10) discute outros pontos que precisariam ser melhor endereçados na cláusula 4(3): em primeiro lugar, o fato de que o opt-out só se aplica a reproduções e extrações posteriores ao recebimento da cláusula – como já discutido, é possível

¹⁷⁴ “(18) (...) No caso de conteúdos que tenham sido publicamente disponibilizados em linha, apenas se deverá considerar adequado reservar esses direitos mediante a utilização de meios de leitura ótica, incluindo metadados e condições gerais de um sítio Internet ou de um serviço. As outras utilizações não deverão ser afetadas pela reserva de direitos para efeitos de prospeção de textos e dados. Noutros casos, poderá ser apropriado reservar os direitos por outros meios, seja por acordos contratuais ou por uma declaração unilateral.” (União Europeia, 2019).

¹⁷⁵ “(18) Esta exceção ou limitação não deverá, de modo algum, prejudicar a exceção obrigatória aplicável à prospeção de textos e dados para fins de investigação científica prevista na presente diretiva, nem a exceção em vigor prevista para os atos de reprodução temporária previstos no artigo 5.º, n.º 1, da Diretiva 2001/29/CE” (União Europeia, 2019).

remover um item de um repositório, mas é mais difícil “destreinar” uma IA já treinada. Em segundo lugar, múltiplas versões de uma mesma obra podem estar contidas em um modelo ou repositório – ou mesmo adaptações feitas por outros artistas (Guadamuz, 2024) – impondo a desenvolvedores a responsabilidade de verificar objeto por objeto dentro das fontes usadas. Para o autor, isto é dificultado ainda mais pelo fato de que mídias diferentes possuem especificidades que tornam este processo mais complexo, como fragmentos de textos e arquivos em áudio e vídeo. Finalmente, a cláusula ignora a possibilidade do desenvolvimento de sistemas que podem se reverter em usos benéficos para os titulares, como ferramentas de detecção de plágio. Tratam-se de preocupações que, majoritariamente, refletem as mesmas considerações de Margoni (2024) sobre o acesso lícito e o potencial destes dispositivos em impor exigências impossíveis aos usuários, tornando os Artigos 3 e 4 efetivamente inócuos.

Em face das observações acima, alguns comentaristas apontam que, embora as limitações nos artigos 3 e 4 tragam diversos pontos positivos (tais como a imposição de adoção obrigatória, a não-diferenciação de usuários ou de finalidades) o texto dos dispositivos ainda se encontra demasiadamente restrito em diversos aspectos (em relação, por exemplo, aos atos contemplados e à vulnerabilidade das limitações a barreiras contratuais e tecnológicas), gerando precedentes perigosos para atividades de mineração de textos e dados que, em tese, seriam lícitas, mas não se encaixam nesta limitação. Isto levaria a obstáculos para a inovação no campo da atuação empresarial e a uma posição competitiva desvantajosa da União Europeia no desenvolvimento de tecnologias centradas em IA (Geiger, Frosio e Bulayenko, 2018; Quintais, 2020, Ducato e Strowel, 2021; Geiger, 2021; Margoni e Kretschmer, 2022; Keller, 2023; Tyagi, 2023).

Mais recentemente, os desafios impostos pela Inteligência Artificial levaram à instituição da Regulação (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024: o *Artificial Intelligence Act*, ou *AI Act* (União Europeia, 2024). No documento, a menção aos direitos de autor é feita nos itens 104 a 109 do preâmbulo, que visam resguardar modelos de IA abertos, ao mesmo tempo em que reafirmam a reserva legal do art. 4 da Diretiva CDSM. Aqui, cabe afirmar que, do ponto de vista do Direito Autoral, a regulação se restringe a modelos de IA de finalidade geral, que excluem expressamente aqueles desenvolvidos para pesquisa ou desenvolvimento de protótipos.¹⁷⁶ Assim, o treinamento de inteligências artificiais fora dessas

¹⁷⁶ “Artigo 3.º

Definições

Para efeitos do presente regulamento, entende-se por:

(...)

exceções, para além das obrigações impostas pelas Diretivas já mencionadas, passa a estar sujeito também a requisitos de transparência que envolvem a criação, manutenção e divulgação, nos termos da Regulação, de uma documentação a respeito do conteúdo utilizado.¹⁷⁷

Ademais, nota-se que o item 106 do preâmbulo inclusive destaca que estas regras se aplicam a quaisquer desenvolvedores ou empresas atuantes em solo europeu, independentemente de o treinamento ter ocorrido fora do bloco – uma medida que visa resguardar desenvolvedores europeus de efeitos concorrenciais adversos por parte de sistemas de IA sediados fora da União Europeia. Trata-se de restrições que não interferem diretamente com as já complexas regulações existentes até então, mas que adicionam mais encargos para a mineração de textos e dados para pesquisadores não afiliados a instituições.

3.2.4. O Reino Unido e o sistema “híbrido” do *fair dealing*

Anteriormente, observou-se que o modelo de cláusula aberta praticado pelos EUA era, na verdade, um modelo híbrido, que também contava com uma lista de usos livres em âmbito estatutário. Para além disso, apontou-se que a criação de usos livres pela via judicial não obstava a promulgação de instrumentos estatutários similares para a mineração de textos e dados – tal

«Modelo de IA de finalidade geral», um modelo de IA, inclusive se for treinado com uma grande quantidade de dados utilizando a autossupervisão em escala, que apresenta uma generalidade significativa e é capaz de executar de forma competente uma vasta gama de tarefas distintas, independentemente da forma como o modelo é colocado no mercado, e que pode ser integrado numa variedade de sistemas ou aplicações a jusante, exceto os modelos de IA que são utilizados para atividades de investigação, desenvolvimento ou criação de protótipos antes de serem colocados no mercado;” (União Europeia, 2024).

¹⁷⁷ “Artigo 53.º

Obrigações dos prestadores de modelos de IA de finalidade geral

1. Os prestadores de modelos de IA de finalidade geral devem:

(...)

b) Elaborar, manter atualizadas e disponibilizar informações e documentação aos prestadores de sistemas de IA que pretendam integrar o modelo de IA de finalidade geral nos seus sistemas de IA. Sem prejuízo da necessidade de respeitar e proteger os direitos de propriedade intelectual e as informações comerciais de caráter confidencial ou segredos comerciais, em conformidade com o direito da União e o direito nacional, as informações e documentação devem:

i) permitir que os prestadores de sistemas de IA tenham uma boa compreensão das capacidades e limitações do modelo de IA de finalidade geral e cumpram as suas obrigações nos termos do presente regulamento, e

ii) conter, no mínimo, os elementos previstos no anexo XII;

c) Aplicar uma política para dar cumprimento ao direito da União em matéria de direitos de autor e direitos conexos e, em especial, identificar e cumprir, nomeadamente através de tecnologias de ponta, uma reserva de direitos expressa nos termos do artigo 4.º, n.º 3, da Diretiva (UE) 2019/790;

d) Elaborar e disponibilizar ao público um resumo suficientemente pormenorizado sobre os conteúdos utilizados para o treino do modelo de IA de finalidade geral, de acordo com um modelo facultado pelo Serviço para a IA.

2. As obrigações estabelecidas no n.º 1, alíneas a) e b), não se aplicam aos prestadores de modelos de IA lançados ao abrigo de uma licença gratuita e aberta que permita o acesso, a utilização, a modificação e a distribuição do modelo, e cujos parâmetros, incluindo as ponderações, as informações sobre a arquitetura do modelo e as informações sobre a utilização do modelo, sejam disponibilizados ao público. Esta exceção não se aplica a modelos de IA de finalidade geral com riscos sistémicos.” (União Europeia, 2024).

como observado em Singapura, que ainda figura como o único país a lançar mão desta possibilidade de forma direta.

Contudo, merece destaque o caso do Reino Unido, que adota um modelo híbrido distinto dos demais: a *negociação justa*, ou *fair dealing*, também seguido por países como Canadá e outras ex-colônias inglesas. Este trabalho se concentrará sobre a legislação britânica não apenas por esta já haver implementado um uso livre explícito para mineração de textos e dados neste sistema, mas também como exemplo do impacto dos debates sobre IA Generativa sobre este tipo de regulação.

Neste sentido, embora seu nome original evoque a ideia do *fair use*, o *fair dealing* funciona de forma marcadamente diferente e mais restritiva: um determinado uso (em um caso concreto) se enquadre em uma das hipóteses de uso livre listadas em lei (artigos 28 a 76 na lei britânica). Em caso positivo, ainda caberá a um tribunal avaliar se esta aplicação em particular das hipóteses foi *justa*, seguindo um teste aberto que pode considerar elementos similares aos do uso justo estadunidense, tais como o propósito do uso, a forma de acesso à obra, o quanto esta foi usada e qual o benefício para o interesse público. Em outras palavras, enquanto o *fair use* determina que o Legislativo ou o Judiciário participam *alternativamente* na construção das L&E, o *fair dealing* impõe que legisladores e cortes devem participar *simultaneamente* deste processo (Geist, 2024).

Feitas estas considerações, observa-se que a lei de Direito Autoral do Reino Unido prevê, em seu artigo 29A, uma possibilidade de reprodução livre para fins de análise de dados e texto, desde que para fins não-comerciais, em moldes similares aos da Diretiva 96/9/CE – com a diferença de que, ao contrário da regulação da UE, não se permite o estabelecimento de cláusula contratual que restrinja a aplicação desse dispositivo (Reino Unido, 2023).¹⁷⁸ Outro

¹⁷⁸ “**29A Copies for text and data analysis for non-commercial research.**

(1)The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—

(a)the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and

(b)the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).

(2)Where a copy of a work has been made under this section, copyright in the work is infringed if—

(a)the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or

(b)the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.

(3)If a copy made under this section is subsequently dealt with—

(a)it is to be treated as an infringing copy for the purposes of that dealing, and

(b)if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.

(4)In subsection (3) “dealt with” means sold or let for hire, or offered or exposed for sale or hire.

ponto de destaque é seu artigo 50D, que permite o acesso a uma base de dados mediante quaisquer medidas consideradas necessárias, desde que o uso pretendido seja permitido na lei (Reino Unido, 2023).¹⁷⁹

Mais recentemente, o governo britânico cogitou expandir esta limitação para incluir usos comerciais, com a intenção de evitar que os custos de licenciamento criassem uma barreira para desenvolvedores, pesquisadores, jornalistas e outros segmentos da sociedade. Diante das possibilidades elencadas em uma consulta pública conduzida em 2022, o governo originalmente optou por seguir adiante com a expansão, sem a possibilidade de *opt-out* (como no art. 4 da Diretiva CDSM).¹⁸⁰ Entretanto, o governo posteriormente sinalizou um recuo em sua decisão, mantendo sua legislação inalterada até o momento. Tal decisão parece encontrar motivação não apenas em a pressões de titulares – como a Publishers Association, que arguiu que o licenciamento para mineração de textos e dados constituiria hoje uma modalidade normal de exploração de obras no mercado editorial britânico –¹⁸¹, como também diante de crescentes preocupações com a aplicação da mineração de textos e dados no treinamento de IA Generativa (Reino Unido, 2022b; 2022c).¹⁸²

3.3. Limitações para mineração de textos e dados na América Latina

(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.” (REINO UNIDO, 2023).

¹⁷⁹ “50D Acts permitted in relation to databases.

(1) It is not an infringement of copyright in a database for a person who has a right to use the database or any part of the database, (whether under a licence to do any of the acts restricted by the copyright in the database or otherwise) to do, in the exercise of that right, anything which is necessary for the purposes of access to and use of the contents of the database or of that part of the database.

(2) Where an act which would otherwise infringe copyright in a database is permitted under this section, it is irrelevant whether or not there exists any term or condition in any agreement which purports to prohibit or restrict the act (such terms being, by virtue of section 296B, void)” (Reino Unido, 2023).

¹⁸⁰ “58. The Government has decided to introduce a new copyright and database right exception which allows TDM for any purpose. The Government will identify suitable legislation to make the required changes in due course.

59. Introducing an exception which applies to commercial TDM will bring benefits to a wide range of stakeholders in the UK. These include researchers, AI developers, small businesses, cultural heritage institutions, journalists, and engaged citizens. Targeted products and services will benefit businesses and customers. Research outcomes could also benefit the wider public. (...) The benefits will be reducing the time needed to obtain permission from multiple rights holders and no licence fee to pay. This will speed up the TDM process and development of AI.

(...)

61. Rights holders will no longer be able to charge for UK licences for TDM and will not be able to contract or opt-out of the exception (...). However, rights holders will still have safeguards to protect their content. The main safeguard will be the requirement for lawful access.” (Reino Unido, 2022).

¹⁸¹ Para mais informações, ver: <https://www.publishers.org.uk/wp-content/uploads/2022/08/22-8-Briefing-note-for-IPO-on-TDM.pdf>. Acesso em 31 ago. 2024.

¹⁸² “The recent Intellectual Property Office consultation on artificial intelligence and intellectual property sought evidence and views on text and data mining. A response was published in June. We recognise that the creative industries have significant concerns about the potential impacts of the TDM proposal and as a result, we are reflecting on whether to progress it in its current form.” (Reino Unido, 2022c).

Por fim, feitas as considerações acima, convém observar de que maneira este debate regulatório tem alcançado os países latino-americanos. Mais especificamente, cumpre averiguar se há regulações sobre o tema da mineração de textos e dados no Direito Autoral e, em caso positivo, quais modelos foram implementados ou se pretendem adotar.

Nesse sentido, o mapa de Flynn et al (2022) sugere um cenário de pouca abertura nesta região: a maioria dos países da região possui regimes de L&E bastante restritivos, especialmente na América do Sul, que figura no mapa dos autores como o local com o maior número de países em vermelho – onde, portanto, não há margem legal para o uso de obras na análise para fins de pesquisa. De forma similar, a Aliança da Sociedade Civil Latino-Americana para o Acesso Justo ao Conhecimento (2023, p. 5,7) elencou as legislações de todos os países latino-americanos e constatou que, à época, praticamente nenhuma regulava a mineração de textos e dados de forma direta e poucas previam hipóteses legais que poderiam ser indiretamente aproveitadas para essas técnicas – e, mesmo assim, com diversas restrições. Assim, Charquero e Rangel (2023) notam que o abismo regulatório apontado por Flynn e Palmedo (2017), portanto, parece demonstrar uma continuidade especialmente marcante nesta região, levando a preocupações sobre as dificuldades trazidas por este ambiente regulatório a pesquisadores locais (Bertón, 2021; Botero, 2024). Nesse sentido, Schirru et al (2024) apontam que diversos agentes na América Latina expressaram preocupação com a necessidade de estabelecer mais garantias para atividades de pesquisa intensiva em dados.

Entretanto, os últimos anos começaram a ver movimentos na região em resposta a esta demanda, a fim de retificar o atual cenário de incerteza regulatória. Nesse sentido, o primeiro país latino-americano a incorporar instrumentos legais aplicáveis à mineração de textos e dados foi o Equador (Charquero, 2021), que possui uma regra (art. 212 num. 9 viii do *Código Ingenios*)¹⁸³ bastante restrita: somente protege bibliotecas e arquivos pela atividade de mineração de textos empreendida pelos seus usuários, quando estes houverem agido de boa-fé. No entanto, observa-se que, para além desta limitação, a legislação equatoriana também possui um modelo misto com cláusula aberta prevista em lei. Desta forma, há margem para usos livres mais abertos

¹⁸³ “La minería de textos. Las bibliotecas o archivos y sus funcionarios estarán exentos de responsabilidad por los actos que realicen sus usuarios siempre y cuando actúen de buena fe y tengan motivos razonables para creer que la obra protegida por derechos de autor o la prestación protegida por derechos conexos se ha utilizado en el marco permitido por las limitaciones y excepciones previstas en el presente Parágrafo o de un modo que no está restringido por los derechos sobre la obra o prestación, o que dicha obra o prestación se encuentra en el dominio público o bajo una licencia que permita su uso” (Equador, 2016).

que beneficiem o uso de técnicas de coleta e análise de dados na pesquisa, bastando que sejam produzidas decisões judiciais a respeito no país.

Ao mesmo tempo, já se discute em outros países da região o papel da mineração de textos e dados na pesquisa e a necessidade de se estabelecer marcos regulatórios de maneira similar à vista em países do Norte Global. No México (2020), por exemplo, já se propôs, em 2020, a implementação não apenas de uma cláusula aberta,¹⁸⁴ mas também de uma limitação de Direito Autoral que se dirija simplesmente ao direito de reprodução e extração de obras licitamente acessadas para fins específicos de mineração de textos e dados (México, 2020).¹⁸⁵

Em 2023, Charquero e Rangel (2023, p. 21-24) constataram que sete países na América Latina já haviam criado grupos de trabalho e consultas públicas entre 2018 e 2022 referentes à inteligência artificial. Dentre estes, cinco mencionavam explicitamente a regulação de propriedade intelectual como um componente relevante – com Chile e Argentina dando atenção mais direta à proteção de bases de dados e apenas o Brasil (como se verá mais adiante) mencionando explicitamente a necessidade de se discutir uma limitação de Direito Autoral para a pesquisa intensiva em dados.

Mais recentemente, estas discussões começaram a tomar a forma de novas propostas legislativas: No Uruguai, a Comisión de Educación y Cultura propôs uma limitação simples que permite a reprodução para análise computacional com fins de pesquisa não-comercial.¹⁸⁶ Já no Chile, o projeto de lei atualmente em tramitação para regular a IA no país traz, consigo, um uso livre bem mais amplo e detalhado em moldes similares aos vistos no Japão: prevê a reprodução, adaptação, distribuição e comunicação ao público no processo de análise estatística de grandes volumes de dados que não constituam uma exploração velada das obras.¹⁸⁷

¹⁸⁴ O texto completo está disponível em: https://www.senado.gob.mx/65/gaceta_del_senado/documento/109633. Acesso em 4 set. 2024.

¹⁸⁵ “Artículo 148.- El uso justo de obras literarias y artísticas, sin autorización del titular del derecho patrimonial y sin remuneración, para fines como la crítica, el comentario, la difusión de noticias, la enseñanza, la educación o la investigación no constituye infracción a los derechos de autor.

(...)

De manera enunciativa más no limitativa, se considerará como uso justo:

(...)

IX. La reproducción y extracción con el fin de realizar la minería de textos y datos de obras a las que tengan acceso lícito.” (México, 2020).

¹⁸⁶ “Artículo 3º.- Agréganse al artículo 45 de la Ley N° 9.739, de 17 de diciembre de 1937, los siguientes numerales: (...)

14) Son igualmente lícitas: (...)

b) La reproducción de obras con fines de análisis computacional, siempre que se realicen en el marco de la investigación no comercial” (Uruguai, 2024).

¹⁸⁷ “Artículo 31.- Incorpórase en la ley N° 17.336 sobre Propiedad Intelectual el siguiente artículo 71 T, nuevo: ‘Artículo 71 T.- Es lícito, sin remunerar ni obtener autorización del titular, todo acto de reproducción, adaptación, distribución o comunicación al público, de una obra licitamente publicada, cuando dicho acto se realice

Ao fim e ao cabo, observa-se um cenário tecnológico e regulatório a nível internacional que começa a influenciar os debates sobre a mineração de textos e dados na América Latina, não podendo ser ignorado quando analisarmos de que maneira o ordenamento jurídico nacional acomoda e adapta esta discussão às suas especificidades.

Qualquer caminho para a resolução deste problema, contudo, também deve levar em consideração os próprios anseios e práticas da comunidade científica no país. Em outras palavras, debater uma regulação para a mineração de textos e dados para fins de investigação no Brasil perpassa compreender a posição da academia dentro do ecossistema nacional de inovação, bem como compreender de que forma essas técnicas surgem na atividade acadêmica e, principalmente, como a gestão e apropriação dos materiais de pesquisa é vista entre aqueles que participam da produção do conhecimento. É o que veremos a seguir.

exclusivamente para la extracción, comparación, clasificación, o cualquier otro análisis estadístico de datos de lenguaje, sonido o imagen, o de otros elementos de los que se componen un gran número de obras o un gran volumen de datos, siempre que dicha utilización no constituya una explotación encubierta de la obra o de las obras protegidas.” (Chile, 2024).

CAPÍTULO 4: MINERAÇÃO DE TEXTOS E DADOS NO ECOSISTEMA DE PESQUISA

A construção de uma regulação destinada ao uso de obras protegidas em atividades de pesquisa necessita de um debate com ampla participação dos setores interessados (Alvarenga, 2019). Isto requer que se observe o ponto de vista não apenas de quem, à primeira vista, tem o escopo de seus direitos reduzido – ou seja, artistas e titulares –, mas também o daqueles que serão (ou pretende-se que sejam) os principais usuários desse campo: os próprios envolvidos com a atividade de pesquisa.

Assim, este capítulo se dedica a observar em que estado se encontra o debate, acerca das questões referentes ao uso de obras para atividades de pesquisa e de que forma a mineração de textos e dados estaria aí inserida – com enfoque sobre a atividade acadêmica enquanto agente de destaque na literatura sobre inovação, embora se ressalve que a regulação sobre as técnicas de coleta e análise de dados também afeta a pesquisa de diversos outros profissionais, como no jornalismo e na conservação. Nesse sentido, portanto, a presente análise mobiliza em especial o referencial teórico referente às práticas acadêmicas e suas visões sobre acesso a dados e propriedade intelectual, combinadas com análise documental e de um conjunto de entrevistas com acadêmicos, representantes de agências de fomento e de entidades de preservação, entre outros. Este debate se inicia com uma consideração acerca do papel da atividade científica nas políticas de inovação, sobretudo brasileiras, para então se debruçar sobre as diversas questões e contradições envolvendo movimentos de abertura de dados e o desejo de protegê-los.

4.1. A interseção entre academia e mercado

A inserção da ciência enquanto alvo de políticas públicas de inovação não é um fenômeno recente. Desde pelo menos o término da Segunda Guerra Mundial, já havia preocupação nos Estados Unidos em relação não apenas à importância da atividade acadêmica na sociedade, mas ao papel que o Estado deveria assumir para estimular a sua continuidade. Em sua comunicação ao governo norte-americano, Bush (1960) destacava a importância de uma política nacional de fomento à pesquisa – seja de forma direta, através de instituições governamentais, seja indiretamente, através de estímulos a entidades privadas.¹⁸⁸

¹⁸⁸ “Advances in science when put to practical use mean more jobs, higher wages, shorter hours, more abundant crops, more leisure for recreation, for study, for learning how to live without the deadening drudgery which has been the burden of the common man for ages past. Advances in science will also bring higher standards of living, will lead to the prevention or cure of diseases, will promote conservation of our limited national resources, and will assure means of defense against aggression. But to achieve these objectives to secure a high level of employment, to maintain a position of world leadership—the flow of new scientific knowledge must be both continuous

Estas iniciativas também se estendem à própria gestão da propriedade intelectual: um exemplo, ainda nos EUA foi o Bayh-Dole Act, uma lei criada para permitir às universidades a possibilidade de patentear e comercializar invenções criadas no interior das mesmas – mesmo que oriundas de financiamento público. Para Marzano (2011, p. 104-106), esta iniciativa teve como objetivo fortalecer redes de conhecimento entre instituições de pesquisa e o setor privado – o que aparecia como uma solução para a erosão do financiamento direcionado aos laboratórios públicos e às pequenas empresas na década de 1970 – resultado da crise econômica oriunda dos choques do petróleo no final da década de 1970 e do crescimento de políticas neoliberais do governo norte-americano à época.

Este movimento de aproximação entre ciência e mercado segue se desdobrando de diversas maneiras – inclusive alterando a lógica de trabalho de universidades. Na década de 1980, Etzkowitz (1983) notou, nos EUA, um movimento de pesquisadores no sentido de aproveitar sua pesquisa para fins comerciais, a fim de reduzir sua dependência do financiamento de agências governamentais ou doações privadas – bem como forma de aumentar a renda para além dos salários e bolsas oferecidos pelas universidades. Isto incluía segundo o autor, a criação de empresas próprias ou a busca de parcerias com empresas já existentes, o que levou a algumas controvérsias sobre a necessidade de balancear o compartilhamento público de informação científica e a proteção comercial dos produtos gerados a partir desta, conforme se verá na seção seguinte.

O crescimento do papel do conhecimento na atividade econômica, manifestado em indústrias como a farmacêutica e informática, levou ao fortalecimento do papel de instituições de pesquisa na literatura sobre inovação: para além de dinâmicas centralizadas no Estado ou com atores completamente autônomos, cresceu a ideia de uma estrutura de inovação em formato de rede, onde parcerias entre centros de pesquisa, firmas e governo teriam papel fundamental na criação de tecnologia de ponta (Etzkowitz e Leydesdorff, 1995; 2000).¹⁸⁹ Este modo de pensar traz, em si, a ideia de que a produção de conhecimento e a de inovação estão

and substantial (...). From early days the Government has taken an active interest in scientific matters (...). Government scientific agencies have splendid records of achievement, but they are limited in function (...). The publicly and privately supported colleges, universities, and research institutes are the centers of basic research. They are the wellsprings of knowledge and understanding. As long as they are vigorous and healthy and their scientists are free to pursue the truth wherever it may lead, there will be a flow of new scientific knowledge to those who can apply it to practical problems in Government, in industry, or elsewhere.” (BUSH, 1960, p. 10-12).

¹⁸⁹ Para autores mais concentrados sobre o papel do Estado na inovação, os agentes públicos atuariam nesse processo seja como facilitadores – o que inclui criar instituições para o processo de geração e difusão de novas tecnologias, o estímulo à competição e propagação de melhores práticas (Ferraz, Paula e Kupfer, 2002) –, seja através de investimentos mais radicais e da imposição de uma visão e missão claras de desenvolvimento para as políticas de inovação (Mazzucato, 2013, p. 22).

intimamente entrelaçadas, em múltiplos níveis, através de fluxos de informação a nível local, nacional e global, de forma física ou não (Carayannis e Campbell, 2009, p. 215-216).¹⁹⁰ Ademais, a comunicação e a mídia também ocupam um papel nesse processo, difundindo informação e cultura e influenciando a opinião pública – o que, por sua vez, influencia a política (Carayannis e Campbell, 2009; Carayannis e Rakhmatulin, 2014).

Jain, George e Maltarich (2009) observam que este raciocínio voltado para o fomento da participação da pesquisa no processo de inovação e na atividade econômica também se estende aos pesquisadores: entre os profissionais analisados pelos autores, a PI também surge não simplesmente como uma ferramenta de ganho financeiro pessoal pela atividade empresarial, mas também como instrumento de captação de recursos e de controle sobre os usos dos resultados da pesquisa – uma preocupação que perdura mesmo em ambientes de compartilhamento mais amplo, como se verá mais adiante.

Assim, surge a noção de universidade empreendedora,¹⁹¹ que situa as instituições formais de pesquisa científica em um cenário maior de inovação, no qual desempenham uma variedade de papéis: enquanto fonte de conhecimento e tecnologia (via profissionais qualificados, publicações, consultoria); enquanto ponto de partida para a criação de empresas spin-offs ou como um agente que compõe as redes de inovação, através de interações diretas das instituições de pesquisa com a iniciativa privada e o Estado (por exemplo, através do uso e gestão da propriedade intelectual sobre a pesquisa, de negociações com empresários e apoio técnico a governos) (Guerrero et al, 2016; Liu e Van der Sijde, 2021). Cada universidade, centro de pesquisa e pesquisador manifestará essas características focando em diferentes práticas: seja a produção científica, a criação de empresas e acordos com firmas já existentes ou o maior envolvimento na construção de políticas públicas (Oliveira e Rodrigues, 2022).

O Brasil não escapou desta tendência, incluindo a universidade como um elo importante dentro de sua política de inovação, especialmente no que diz respeito à transformação do conhecimento científico em um ativo econômico para o país. Um dos instrumentos regulatórios

¹⁹⁰ Esta noção sistêmica do ecossistema de inovação pode ser vista também em outros trabalhos que não necessariamente têm enfoque na atividade de pesquisa: por exemplo, Malerba (2002), ao tratar dos sistemas setoriais de inovação, aponta para a formação de agrupamentos setoriais que compartilham bases de conhecimento, tecnologias, interações entre agentes e instituições específicas.

¹⁹¹ E, paralelamente, da “firma acadêmica”, que ressalta a importância das empresas (seja no todo ou em setores internos específicos) buscarem maior diálogo com centros de pesquisa, unir pesquisa básica e aplicada, estimular a codificação do conhecimento gerado pelos funcionários, criar redes de colaboração, entre outras iniciativas. O propósito desse movimento é agregar valor à pesquisa e desenvolvimento no âmbito da empresa através do apoio intelectual que as universidades podem oferecer (Campbell e Güttel, 2005, p.171; Carayannis e Campbell, 2009, p. 211-212).

que foram empregados nesse sentido foi a Lei no. 10.973/2004 (a Lei de Inovação), atualizada em 2016, que, dentre diversas disposições, atribuiu às universidades e outros institutos públicos de ciência e tecnologia diversas responsabilidades e prerrogativas no sentido de construir um ambiente institucional interno de incentivo à inovação. Isto inclui não apenas incentivar a participação em empresas, como também a criação de ambientes de criação (ou incubadoras) de empresas, a criação de escritórios de transferência de tecnologia e, tal como o Bayh-Dole Act, a apropriação dos resultados de pesquisas internas via propriedade intelectual (Brasil, 2004; 2016).¹⁹²

Um exemplo prático da concretização desse processo pode ser visto em Renault et al (2011), a respeito da criação de instituições de apoio à inovação a partir da Coordenação de Programas de Pós-Graduação em Engenharia (COPPE) da Universidade Federal do Rio de Janeiro (UFRJ) na década de 1990. A partir deste cenário, o estudo acompanha quatro empresas de base tecnológica fundadas por professores e alunos – notando-se que duas delas possuem material proprietário (um dos quais, inclusive, licenciado para outra empresa), enquanto outra mantém sua tecnologia sob sigilo. Note-se que, em todos os casos observados, trata-se de produtos e serviços originados de projetos de pesquisa realizados dentro do ambiente acadêmico: usos, em ambiente comercial, de tecnologias desenvolvidas no seio de instituições sem fins lucrativos. Um dos mecanismos institucionais que foram empregados nesse processo foi a atuação do escritório de transferência de tecnologia, ou Núcleo de Inovação Tecnológica (NIT): entidade ou órgão interno à universidade designado pela Lei de Inovação como o responsável, dentro da universidade, por gerenciar os direitos de propriedade intelectual sobre a produção

¹⁹² “Art. 15-A. A ICT de direito público deverá instituir sua política de inovação, dispondo sobre a organização e a gestão dos processos que orientam a transferência de tecnologia e a geração de inovação no ambiente produtivo, em consonância com as prioridades da política nacional de ciência, tecnologia e inovação e com a política industrial e tecnológica nacional. (Incluído pela Lei nº 13.243, de 2016)

Parágrafo único. A política a que se refere o **caput** deverá estabelecer diretrizes e objetivos: (Incluído pela Lei nº 13.243, de 2016)

I - estratégicos de atuação institucional no ambiente produtivo local, regional ou nacional; (Incluído pela Lei nº 13.243, de 2016)

II - de empreendedorismo, de gestão de incubadoras e de participação no capital social de empresas; (Incluído pela Lei nº 13.243, de 2016)

III - para extensão tecnológica e prestação de serviços técnicos; (Incluído pela Lei nº 13.243, de 2016)

IV - para compartilhamento e permissão de uso por terceiros de seus laboratórios, equipamentos, recursos humanos e capital intelectual; (Incluído pela Lei nº 13.243, de 2016)

V - de gestão da propriedade intelectual e de transferência de tecnologia; (Incluído pela Lei nº 13.243, de 2016)

VI - para institucionalização e gestão do Núcleo de Inovação Tecnológica; (Incluído pela Lei nº 13.243, de 2016)

VII - para orientação das ações institucionais de capacitação de recursos humanos em empreendedorismo, gestão da inovação, transferência de tecnologia e propriedade intelectual; (Incluído pela Lei nº 13.243, de 2016)

VIII - para estabelecimento de parcerias para desenvolvimento de tecnologias com inventores independentes, empresas e outras entidades. (Incluído pela Lei nº 13.243, de 2016)” (Brasil, 2004).

gerada dentro da instituição, bem como a negociação e transferência destes para os pesquisadores ou para empresas externas à universidade (Brasil, 2004).¹⁹³¹⁹⁴

Santos e Torkomian (2013) observaram a atuação dos NITs no cenário brasileiro e observam um retorno geralmente positivo desta política, apontando para o aumento no registro de patentes e de licenciamentos com empresas, além do aumento progressivo na receita dos institutos públicos de pesquisa resultante dos *royalties* gerados com os contratos de transferência. As pesquisadoras reconheceram, porém, que estes ganhos ainda se demonstraram relativamente modestos à época, ainda que promissores.

Esta última observação também parece sugerir um cenário no qual ainda há trabalho a ser feito para aumentar a aproximação entre universidades e empresas: os trabalhos de Fischer, Schaeffer e Silveira (2018) e de Fischer, Schaeffer e Vonortas (2020) sugerem, por exemplo, que a localização de universidades de alta qualidade não parece exercer muita influência sobre a atração de investimentos em atividades produtivas intensivas em conhecimento e que as instituições de pesquisa parecem ainda ocupar um papel secundário entre os potenciais parceiros de empresas, embora tenha-se notado um gradual aumento em co-patenteamentos e de aumento em capacidades de inovação dentro das firmas.

Estas observações parecem coincidir com o cenário mostrado na última Pesquisa de Inovação (PINTEC) Semestral de 2022 do IBGE, relativa à atividade industrial, de todas as

¹⁹³ “Art. 11. Nos casos e condições definidos em normas da ICT e nos termos da legislação pertinente, a ICT poderá ceder seus direitos sobre a criação, mediante manifestação expressa e motivada e a título não oneroso, ao criador, para que os exerça em seu próprio nome e sob sua inteira responsabilidade, ou a terceiro, mediante remuneração. (Redação pela Lei nº 13.243, de 2016)” (Brasil, 2004).

Parágrafo único. A manifestação prevista no caput deste artigo deverá ser proferida pelo órgão ou autoridade máxima da instituição, ouvido o núcleo de inovação tecnológica, no prazo fixado em regulamento.

¹⁹⁴ “Art. 16. Para apoiar a gestão de sua política de inovação, a ICT pública deverá dispor de Núcleo de Inovação Tecnológica, próprio ou em associação com outras ICTs. (Redação pela Lei nº 13.243, de 2016)

§ 1º São competências do Núcleo de Inovação Tecnológica a que se refere o **caput**, entre outras: (Redação pela Lei nº 13.243, de 2016)

I - zelar pela manutenção da política institucional de estímulo à proteção das criações, licenciamento, inovação e outras formas de transferência de tecnologia;

(...)

IV - opinar pela conveniência e promover a proteção das criações desenvolvidas na instituição;

V - opinar quanto à conveniência de divulgação das criações desenvolvidas na instituição, passíveis de proteção intelectual;

VI - acompanhar o processamento dos pedidos e a manutenção dos títulos de propriedade intelectual da instituição.

VII - desenvolver estudos de prospecção tecnológica e de inteligência competitiva no campo da propriedade intelectual, de forma a orientar as ações de inovação da ICT; (Incluído pela Lei nº 13.243, de 2016)

VIII - desenvolver estudos e estratégias para a transferência de inovação gerada pela ICT; (Incluído pela Lei nº 13.243, de 2016)

IX - promover e acompanhar o relacionamento da ICT com empresas, em especial para as atividades previstas nos arts. 6º a 9º; (Incluído pela Lei nº 13.243, de 2016)

X - negociar e gerir os acordos de transferência de tecnologia oriunda da ICT. [\(Incluído pela Lei nº 13.243, de 2016\)](#)” (Brasil, 2004).

empresas analisadas, 25,7% afirmaram manter cooperação com outras organizações dentro da infraestrutura de ciência, tecnologia e inovação – o que, embora represente um aumento em relação ao índice do ano anterior (19,7%), ainda é inferior ao percentual (já relativamente baixo) de relações de cooperação com fornecedores (35,3%), clientes ou consumidores (31,2%) e empresas de consultoria (28,1%) (Instituto Brasileiro de Geografia e Estatística, 2022).

Aqui, observa-se uma clara associação entre a aproximação das universidades e instituições de pesquisa e o mercado, enquanto objetivo (e objeto) de políticas de inovação – com a propriedade intelectual sendo apontada como um elemento intermediador desse processo de geração de valor. Outros autores, contudo, enxergam esta e outras iniciativas de fomento de forma mais reticente: enquanto marcos de um processo de “privatização” e “cerceamento” da pesquisa científica, com chancela e incentivo do Estado, com consequências negativas para a atividade de produção e compartilhamento do conhecimento. Esta seria a base de movimentos reativos dentro da academia em nome de uma ciência aberta e, posteriormente, de dados abertos, que serão objeto da seção a seguir.

4.2. Ciência aberta e compartilhamento de dados na pesquisa

Para diversos autores, tais políticas, especialmente no final do Século XX, também seriam produto de uma pressão crescente, por parte de setores emergentes na fronteira tecnológica de países industrializados, para garantir maior proteção frente a concorrentes no cenário internacional (Argyres e Liebeskind, 1998; Nelson, 2003; Orsi e Coriat, 2006). Isto culminaria na consolidação do que Drahos e Braithwaite (2002) chamam de “feudalismo informacional”: um processo crescente de apropriação de um conhecimento outrora público por mãos privadas através da propriedade intelectual e restrições tecnológicas ao acesso – criando uma estrutura de poder cultural e informacional concentrada em “feudos” sob o controle de agentes privados, que rivalizam com o poder dos Estados – o que é exacerbado pelo processo de globalização, que limita o alcance da regulação estatal.

Por outro lado, no âmbito da atividade científica, Nelson aponta para a necessidade de preservação de um patrimônio científico comunitário (*os scientific commons*),¹⁹⁵ destacando que a ciência possui um caráter cumulativo, onde a realização de testes e pesquisas depende do trabalho oriundo de pesquisas anteriores. Por esta razão, o autor argumenta que ampliar os meios para a apropriação de resultados e mesmo de ferramentas técnicas de investigação, ainda que atraia mais investimentos, leva à imposição de barreiras a trajetórias de pesquisa – seja pela

¹⁹⁵ Para mais informações sobre os diferentes tipos de commons, ver o próximo capítulo.

restrição ao acesso, seja pela multiplicidade de licenças a serem obtidas (Nelson, 2003, p. 22-23).

Os movimentos mais contundentes pela expansão da ciência aberta ganharam força com a Declaração de Budapeste, que preconiza a necessidade de repensar a forma de disponibilizar trabalhos científicos e propõe, para tanto, o compromisso em promover a disseminação de plataformas para depósito de trabalhos pelos próprios autores, bem como o lançamento de mais periódicos de acesso aberto, financiados por formas que não dependam de compras ou assinaturas (Budapeste Open Access Initiative, 2002). Tal movimento foi fortalecido pela Declaração de Berlim sobre o Acesso Aberto ao Conhecimento nas Ciências e nas Humanidades (2003) e do Bethesda Statement on Open Access Publishing (2003).

Um dos desdobramentos desse movimento é a abertura não só do acesso aos resultados da pesquisa, na forma de artigos e textos acadêmicos, mas também de material empírico de suporte – os dados de pesquisa. Surgia, assim, a noção de dados abertos, cujo conceito foi exemplificado em um encontro nos EUA, em 2007, que viria a estabelecer as bases para o Open Government Data – movimento que estabeleceu 8 (oito) princípios para a governança desse tipo de dado:

- 1. Completos:** Todos os dados públicos são disponibilizados. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou privilégio. Embora os recursos de informações não eletrônicas, como artefatos físicos, não estejam sujeitos aos princípios de dados governamentais abertos, é sempre incentivado que esses recursos sejam disponibilizados eletronicamente na medida do possível.
- 2. Primários:** Os dados são coletados na fonte, com o mais alto nível de granularidade possível, não em formas agregadas ou modificadas. Se uma entidade optar por transformar dados por meio de agregação ou transcodificação para uso em um site da Internet criado para usuários finais, ela ainda terá a obrigação de disponibilizar as informações de resolução total em massa para que outras pessoas criem seus próprios sites e preservem os dados para a posteridade.
- 3. Em tempo hábil:** Os dados são disponibilizados tão rapidamente quanto necessário para preservar o valor dos dados.
- 4. Acessíveis:** Os dados estão disponíveis para a mais ampla gama de usuários e para a mais ampla gama de finalidades. Os dados devem ser disponibilizados na Internet de modo a acomodar a mais ampla gama prática de usuários e usos. Isso significa considerar como as escolhas na preparação e publicação de dados afetam o acesso a pessoas com deficiência e como isso pode afetar os usuários de uma variedade de plataformas de software e hardware. Os dados devem ser publicados com protocolos e formatos padrão atuais do setor, bem como protocolos e formatos alternativos quando os padrões do setor impuserem ônus à ampla reutilização dos dados. Os dados não estarão acessíveis se puderem ser recuperados somente por meio da navegação em formulários da Web ou se ferramentas automatizadas não tiverem permissão para acessá-los devido a um arquivo robots.txt, outra política ou restrições tecnológicas.
- 5. Processáveis por máquina:** Os dados são razoavelmente estruturados para permitir o processamento automatizado. Para que os dados possam ser amplamente utilizados, é necessário que eles sejam codificados adequadamente. O texto em formato livre não substitui os registros tabulares e normalizados. Imagens de texto não substituem o texto em si. Documentação suficiente sobre o formato dos dados e os significados dos itens de dados normalizados deve estar disponível para os usuários dos dados.

6. Não-discriminatórios: Os dados estão disponíveis para qualquer pessoa, sem necessidade de registro. O acesso anônimo aos dados deve ser permitido para dados públicos, incluindo o acesso por meio de proxies anônimos. Dados não devem ser escondidos atrás de "jardins murados".

7. Não-proprietários: Os dados estão disponíveis em um formato sobre o qual nenhuma entidade tem controle exclusivo. Os formatos proprietários acrescentam restrições desnecessárias sobre quem pode usar os dados, como eles podem ser usados e compartilhados e se os dados poderão ser usados no futuro. Embora alguns formatos proprietários sejam quase onipresentes, não é aceitável usar apenas formatos proprietários. Da mesma forma, os formatos não proprietários relevantes podem não atingir um público amplo. Nesses casos, pode ser necessário disponibilizar os dados em vários formatos.

8. Livres de licença: Os dados não estão sujeitos a nenhuma regulamentação de direitos autorais, patentes, marcas ou segredos comerciais. Podem ser permitidas restrições razoáveis de privacidade, segurança e privilégio. Como as informações do governo são uma mistura de registros públicos, informações pessoais, obras protegidas por direitos autorais e outros dados não abertos, é importante deixar claro quais dados estão disponíveis e quais licenças, termos de serviço e restrições legais se aplicam. Os dados aos quais não se aplicam restrições devem ser marcados claramente como sendo de domínio público.

(Open Government Data, 2024. Tradução feita pelo autor)¹⁹⁶¹⁹⁷

Embora todos os princípios tenham relevância para este trabalho em maior ou menor grau, é pertinente destacar alguns em especial. Primeiramente, os itens 2 e 5 demonstram uma crescente relevância ante às especificidades técnicas da mineração de textos e dados contemporânea, estas cada vez mais definidas pela crescente depuração e detalhamento dos dados, e por seu entrelaçamento com o desenvolvimento de sistemas de inteligência artificial. Ao mesmo tempo, o item 5 também destaca a necessidade de documentar e formatar adequadamente os

¹⁹⁶ O texto original está disponível em: <https://opengovdata.org>. Acesso em 9 ago. 2024.

¹⁹⁷ A organização ainda elenca 7 princípios adicionais, não elencados pelo grupo de trabalho original. São eles: **“(1) Online e gratuitos:** informação não é significativamente pública se não estiver disponível na Internet gratuitamente ou, pelo menos, não mais do que o custo marginal de reprodução. Ela também deve poder ser encontrada. **(2) Permanentes:** dados devem ser disponibilizados em um local estável na Internet indefinidamente e em um formato de dados estável pelo maior tempo possível. **(3) Confiáveis:** a Recomendação da Association of Computing Machinery sobre Governo Aberto (fevereiro de 2009) afirma que ‘o conteúdo publicado deve ser assinado digitalmente ou incluir atestado de data de publicação/criação, autenticidade e integridade’. As assinaturas digitais ajudam o público a validar a origem dos dados que encontram para que possam confiar que os dados não foram modificados desde que foram publicados. Como a procedência é para documentos publicados originalmente, não é um motivo para impedir que o público modifique documentos governamentais. **(4) Presunção de abertura:** (...) As Diretrizes da Política de Dados Abertos da Sunlight Foundation afirmam: ‘Definir o padrão como aberto significa que o governo e as partes que agem em seu nome disponibilizarão informações públicas de forma proativa e que colocarão essas informações ao alcance do público (on-line), com poucas ou nenhuma barreira para sua reutilização e consumo.... Definir o padrão como aberto significa estar à altura do potencial de nossas informações, analisar o gerenciamento abrangente de informações e tomar decisões que sejam de interesse público’. **(5) Documentados:** A documentação sobre o formato e o significado dos dados ajuda muito a torná-los úteis (...). **(6) Seguros para abrir:** A Recomendação da Association of Computing Machinery sobre Governo Aberto (fevereiro de 2009) declarou: ‘Os órgãos governamentais que publicam dados on-line devem sempre procurar publicar usando formatos de dados que não incluam conteúdo executável’. O conteúdo executável dentro de documentos representa um risco de segurança para os usuários dos dados, pois o conteúdo executável pode ser um malware (vírus, worms etc.). **(7) Desenhados com input do público:** o público está na melhor posição para determinar quais tecnologias de informação serão mais adequadas para os aplicativos que o público pretende criar para si mesmo. Portanto, a contribuição do público é fundamental para a disseminação de informações de forma que elas tenham valor.” (Open Government Data, 2024).

dados – questão essa que aparece como uma dificuldade recorrente na pesquisa, como será visto mais adiante.

Outros princípios que merecem menção mais detalhada são os que mais diretamente abordam questões de propriedade intelectual. O primeiro deles é o item 4, que chama atenção ao se posicionar contrário ao uso de medidas tecnológicas de proteção, enquanto obstáculo para o acesso a dados públicos. Embora não seja o enfoque do manifesto, trata-se de um tópico que ganha contornos mais delicados quando consideramos o uso dessas ferramentas para coibir o treinamento de IAs Generativas – o que, como será visto em um momento posterior, pode impactar o ambiente regulatório sobre o uso de obras na pesquisa.

Ainda no âmbito da PI, outros pontos de destaque são itens 7 e 8, contra a apropriação dos dados públicos via Direito Autoral – opondo-se tanto à concentração dos dados em bases proprietárias como à ideia de que dados possam ser protegidos por Direito Autoral. Entretanto, há aceitação de que obras protegidas possam estar incluídas nos repositórios e que de que haja alguma separação entre estas e os dados. Conceitualmente, é possível interpretar que o texto faz menção a dados puramente dentro de sua conceituação como registros factuais levantada anteriormente (citar), mas é necessário ter em mente a diversidade do que pode ser considerado dado para alguns pesquisadores, o que talvez demande a discussão de um consenso a respeito do que incluir nesta categoria.

Cabe destacar que este projeto, como se pode observar em diversos preceitos acima, apresentava um interesse marcadamente voltado para dados públicos. Contudo, Machado afirma que o conceito se expandiu para outras áreas, incluindo a atividade científica (Machado, 2015, p. 212). No Brasil, esta observação ganha relevância ainda maior quando se tem em mente o papel das instituições públicas de ensino, pesquisa e extensão no país: segundo dados oferecidos pela Clarivate (titular do portal Web of Science) e levantados por especialistas, a esmagadora maioria das publicações científicas produzidas na última década é oriunda de universidades públicas: das 100 instituições que mais submeteram trabalhos, apenas 17 são privadas – com apenas 3 figurando entre as 50 primeiras (Cruz, 2019). Isto levou à declaração do presidente da Academia Brasileira de Ciências, em 2019, de que, considerando a produção científica publicada entre 2011 e 2016 na Web of Science, “mais de 95% das publicações referem-se às universidades públicas, federais e estaduais.” (Academia Brasileira de Ciências, 2019).

Por extensão, é razoável presumir que se trata de um ambiente cuja produção acadêmica é predominantemente financiada por recursos públicos. É evidente, portanto, a relevância de se discutir até que ponto o material de pesquisa gerado no Brasil se caracteriza como dados

públicos, dado o interesse da sociedade em acessar e utilizar uma produção custeada pela mesma.

Mais recentemente, também cresceram demandas por uma maior institucionalização do acesso à pesquisa no âmbito jurídico, sob a forma de um direito humano ao uso e aproveitamento do conhecimento científico, enquanto patrimônio da humanidade. Chapman (2009), por exemplo, pauta a questão como um direito aos benefícios do progresso científico e suas aplicações, não explicitado no Direito Internacional, que geraria, para os países, a obrigação de tomar atitudes no sentido de viabilizar esse acesso. A autora, entretanto, aponta para a dificuldade de se conceber a materialização dessa obrigação de forma universal, especialmente diante de uma possível falta de estrutura material e institucional em países do Sul Global.

A noção de “direito à ciência” enquanto direito de acesso aos produtos da atividade científica encontraria eco em outros trabalhos, que destacam a necessidade de se colocar o conhecimento científico como um bem público – e, portanto, a imprescindibilidade da participação ativa da sociedade em sua construção e no desfrute de seus resultados. Para Chapman e Wyndham (2013), por exemplo, isto deve ocorrer por meio do acesso às aplicações e à produção científica, da possibilidade de produzir ciência e da participação social nas políticas sobre ciência. Shaver (2015b) se posiciona de forma similar, embora reconheça uma certa resistência ao reconhecimento de novos direitos humanos diante do temor de que nasçam diluídos em importância.¹⁹⁸

Mais recentemente, em âmbito mais regional, Oriakhogba (2022) coloca a necessidade de se repensar o Direito Autoral na África, cujos obstáculos para o trabalho de museus, educadores e pesquisadores teriam se tornado evidentes com a pandemia de COVID-19. A despeito das preocupações levantadas acima, o autor toma como base as declarações de direitos de diversos países africanos e identifica a compatibilidade desse direito com os compromissos assumidos por esses Estados a nível nacional, regional e internacional, no sentido de ativamente prover uma infraestrutura adequada para a efetivação desse direito.. Também se observa a compatibilidade dessa garantia com respeito aos interesses dos autores – o que inclui os próprios pesquisadores.

Samtani (2023) se volta para a pesquisa como uma atividade de busca sistemática pelo conhecimento, através da experimentação e investigação de fatos, teorias e modelos. A autora

¹⁹⁸ Em outro trabalho, contudo (Shaver, 2015a), a autora se manifesta pelo que chama de o “direito de ler”, pautado na liberdade (rejeição à censura), capacidade (acesso à educação) e disponibilidade.

localiza o direito de realizar essas atividades de forma implícita no Pacto Internacional dos Direitos Civis e Políticos e no Pacto Internacional dos Direitos Econômicos Sociais e Culturais, logicamente ligado à concretização de diversos direitos culturais, bem como o progresso científico. Tal direito envolveria garantir a disponibilidade, acessibilidade e qualidade da produção de conhecimento – o que inclui liberdade de pesquisa, não-discriminação e manutenção de repositórios de fácil acesso – e só poderia ser limitado por instrumentos legais, de acordo com um exame de proporcionalidade com outros direitos.

Já Beiter (2023) enfatiza a ciência como um bem público global, ligado ao direito ao desenvolvimento e ao dever de solidariedade internacional no que diz respeito à facilitação de transferências de tecnologia que culminem na melhoria da qualidade de vida de países em desenvolvimento. Desta forma, o conhecimento científico não só necessita de validação internacional, como sua concentração em grandes editores desincentiva o estabelecimento de redes de conhecimento locais mais acessíveis para pesquisadores no Sul Global.

Hoje, é possível identificar diversos repositórios que adotam um modelo de acesso aberto. Dois exemplos são o PubMed (Estados Unidos, 2024c), no campo das ciências biológicas e da saúde, e a SciELO, destinada à publicação de periódicos de acesso aberto. Neste último caso, a própria questão da abertura dos dados de pesquisa é mencionada na política de critérios para inclusão na base de dados na plataforma – permitindo aos periódicos incluídos que exijam a abertura dos dados usados pelos pesquisadores, mas demandando que esse tipo de conteúdo, quando pertencente a terceiros, seja devidamente citado e referenciado (SciELO Brasil, 2022).¹⁹⁹

Outro exemplo, desta vez situado no ambiente de instituições públicas de pesquisa, é o Plano de Dados Abertos da Fundação Oswaldo Cruz (2023), cujos objetivos incluem não apenas inventariar os bancos de dados da instituição, mas também avaliar o material de maior relevância ao interesse público e fomentar um ecossistema de abertura e interoperabilidade dos dados governamentais, especialmente na área da saúde (p. 7). Entre os critérios de seleção de repositórios para abertura, nota-se que o plano leva em consideração não apenas quais bases

¹⁹⁹ “A terceira dimensão trata do compartilhamento dos dados, códigos, métodos e outros materiais utilizados e resultantes das pesquisas que geralmente ficam subjacentes nos textos dos artigos publicados pelos periódicos. Esses conteúdos são identificados genericamente como “dados de pesquisa”. Esse compartilhamento é de responsabilidade dos(as) pesquisadores(as) envolvidos na pesquisa e esses devem informar ao periódico na submissão do manuscrito. Aos periódicos cabe, em primeiro lugar, como política editorial e instruções aos autores, promover e posteriormente exigir que os conteúdos subjacentes aos artigos sejam devidamente citados e referenciados. Como prática de ciência aberta, os periódicos podem vir a exigir que, exceto em caso justificados, esses conteúdos na forma de arquivos sejam disponibilizados em acesso aberto em repositórios de dados confiáveis.” (SciELO, 2022, p. 8).

são mais demandadas ou têm informações de disponibilidade obrigatória, mas também o potencial de cada uma delas em fomentar atividades de pesquisa, bem como a geração de negócios (p. 12).

Esta noção de ciência aberta não existe de forma isolada dentro do processo de inovação: ao contrário, ela caminha lado a lado com um crescente movimento de inovação aberta – termo cunhado por Chesbrough para descrever um movimento de maior abertura na produção do conhecimento, potencializado pelo aumento na formação de profissionais qualificados, fontes de financiamento e de oportunidades de parceria com outras entidades – o que envolve uma forma mais colaborativa de pensar a gestão dos bens imateriais produzidos nos centros de pesquisa e desenvolvimento (P&D) empresariais. Isto inclui distanciar-se da necessidade de concentração da propriedade intelectual, fazendo com que empresas usem esquemas de licenciamento para expandir mercados, aprofundar relacionamentos com outros entes na sua rede profissional e aprimorar a expertise de pesquisadores e desenvolvedores internos (Chesbrough, 2003, pp. 43-62).

Partindo-se da noção de que a academia ocupa um papel crucial dentro do ecossistema de inovação, então seria possível presumir uma ligação entre a inovação aberta e a ciência aberta. É o que Gassman e outros autores se propõem a fazer, elencando diversos pontos de interseção entre a atividade acadêmica e a inovação dentro de empresas. Os pesquisadores destacam tendências nesse sentido, tais como o aumento de colaborações e transferências de tecnologias entre universidades e empresas (com iguais direitos de comercializar as aplicações dos trabalhos realizados em conjunto); o movimento de instituições públicas em direção a um foco em ciência aplicada (com áreas de interesse cada vez mais especializadas) e na construção de redes com outros agentes econômicos, buscando múltiplos parceiros e financiadores; a descentralização da pesquisa e terceirização de processos de pesquisa e desenvolvimento (P&D) de empresas para institutos de pesquisa; o aumento de publicações interdisciplinares em formato digital; a criação de infraestruturas e *clusters* de inovação para pesquisa conjunta entre startups e movimentos de doação ou licenciamento gratuito de patentes (Gassman et al, 2011, p. 14-18).

Ao mesmo tempo, esta possível associação entre acesso aberto e inovação aberta também leva a uma preocupação: a de uma adoção acrítica de processos de abertura de dados na ciência leve à sua apropriação indevida por determinados agentes econômicos. Nesse sentido, resgatamos a crítica de Okediji (2018) ao uso da retórica dos direitos humanos dentro do Direito Autoral vista no Capítulo 2 e a estendemos ao direito à pesquisa: de que maneira este se concretizaria dentro de um sistema internacional de propriedade intelectual de forte viés liberal,

pautado em visões de inovação nem sempre pensadas para a infraestrutura e as necessidades de pesquisadores em nações em desenvolvimento ou subdesenvolvidas? Em um ambiente econômico dominado por grandes plataformas sediadas em grandes economias e marcado pela datificação, é importante que se considere, também, o alerta de Gray (2023) de que essa estrutura de mercado ainda é sustentada por relações de apropriação Norte-Sul Global, e o processo de produção e compartilhamento ainda é, em grande parte, pautado por estas relações coloniais. Não se pode, portanto, discutir a apropriação de dados sem que se tenha esta realidade em mente.

Assim, é forçoso notar que a migração de um modelo predominantemente fechado para aberto de compartilhamento de dados pode não ser tão simples: a adoção de um modelo amplo de acesso, mesmo no ambiente acadêmico, ainda conta com uma série de problemas de adesão e implementação – os quais, conforme abordaremos a seguir, guardam estreita relação com a perpetuação destas tensões em relação à apropriação dos dados.

4.3. O compartilhamento de dados por editores e repositórios

Para além das questões inerentes à própria diversidade existente na produção acadêmica, outro desafio para a ciência aberta se manifesta na literatura: a forma pela qual a gestão dos materiais de pesquisa é praticada pelos diversos atores envolvidos no ambiente de investigação científica. Uma das facetas desse problema pode ser vista na atuação das editoras e repositórios científicos.

Nesse sentido, Xalabarder (2023, p. 25-26) observa a consolidação de um “mercado primário” pelo qual bases de dados pertencentes a grandes editoras contendo periódicos e outros materiais são objeto de negócios jurídicos acertados com instituições de ensino e pesquisa, normalmente a preços exorbitantes. A autora nota que a inclusão de práticas de mineração de textos e dados é comumente encontrada neste tipo de contrato, embora aponte-se que a grande variedade de modelos contratuais e de termos de uso, bem como a dificuldade em identificar e localizar titulares, gera confusões para entidades e pesquisadores em relação a quais usos de quais materiais são permitidos. Também se destaca que esta prática, apesar de corriqueira no mercado acadêmico, é pouco vista fora dele: entidades de gestão coletiva de obras artísticas, por exemplo, normalmente não preveem usos para mineração de textos e dados em suas licenças.

Schirru e Margoni (2023), por sua vez, mostram parte dessas relações de forma mais específica, ao observar os termos de uso de dez das maiores editoras acadêmicas do mundo (como Springer, Elsevier e Wiley), notam que todas (com exceção de uma) abordam o uso de

suas publicações em mineração de textos e dados e permitem a prática para fins de pesquisa. Ao mesmo tempo, contudo, são mais restritivas que o observado, por exemplo, na limitação do art. 3 da Diretiva CDSM, ao excluir entidades de preservação. Além disso, as licenças não são claras em relação a quais usos específicos são permitidos: algumas proíbem ou restringem o uso de ferramentas automatizadas, ou exigem que o uso se dê através de uma rede segura ou de sistemas fornecidos pelas próprias editoras – que não são transparentes em relação a como estes mecanismos controlam a disponibilidade das obras utilizadas.

A adoção de plataformas abertas de publicação poderia surgir como alternativa a este padrão, mas esbarra em um problema: a persistência do papel já estabelecido de periódicos fechados em exercer a curadoria e controle de qualidade da produção acadêmica a nível mundial. Mueller-Langer e Scheuffen observam que, embora houvesse um número grande de publicações aderentes ao Directory of Open Access Journals (mais de 8.500), a média de fator de impacto ainda era inferior à de editoras proprietárias, como Elsevier ou Wiley-Blackwell. Esta diferença, segundo os autores, se aprofundaria conforme a área de conhecimento: publicações de acesso aberto figuravam com um destaque maior nas áreas de biológicas e exatas, enquanto o fator de impacto nas áreas de humanas se mostrou bastante escasso (Mueller-Langer e Scheuffen, 2013, p. 5-6).

Ao mesmo tempo, Xalabarder (2023, p. 28-30) também observa que mesmo licenças abertas enfrentam questões próprias para sua implementação, tais como a tensão entre impor garantias de proteção da prioridade no uso dos dados pelo autor de um estudo (sob a forma de direitos de republicação e embargos temporários ao uso por terceiros, por exemplo) e a pressão crescente pelo acesso imediato ao material (comumente implicando na transferência do controle sobre os dados e as publicações para as editoras e instituições), além da falta de financiamento para publicação em periódicos de acesso aberto. Isto se torna um problema quando se observa, como em Solomon e Björk (2012), que muitas editoras proprietárias de alta credibilidade que aderiram a sistemas de acesso aberto ou híbrido comumente financiam estes modelos mediante a cobrança de taxas de processamento de artigos (*Article Processing Charges*, ou APCs). Estas taxas se destinam não aos usuários, mas *aos autores dos estudos*, para que estes tenham seus trabalhos publicados, e geralmente atingem preços extremamente elevados para um pesquisador ou grupo de pesquisadores isolado.

Já entre repositórios, Lin e Strasser (2014) reuniram comentários e sugestões de líderes de administração de dados como uma chamada à ação para que as editoras estabeleçam e apliquem melhores políticas de compartilhamento e documentação de dados. As autoras, então,

chegaram a oito recomendações para editores: 1. Estabelecer e aplicar uma política obrigatória de disponibilidade de dados; 2. contribuir para o estabelecimento de padrões comunitários para o gerenciamento e compartilhamento de dados, bem como 3. a preservação de dados em repositórios confiáveis; 4. fornecer canais formais para o compartilhamento de dados; 5. trabalhar com repositórios para agilizar o envio de dados; 6. exigir a citação apropriada de todos os dados associados a uma publicação – produzidos e usados; 7. desenvolver e reportar indicadores que apoiarão os dados como um produto acadêmico de primeira classe e 8. incentivar o compartilhamento de dados através da promoção do valor dessa prática.

Novamente, observam-se obstáculos por parte de grandes repositórios internacionais. A Clarivate, empresa responsável pelo Web of Science,²⁰⁰ não apenas impõe restrições ao acesso e uso de seu conteúdo por usuários sem afiliação institucional, como também inclui, em seus termos de uso, uma proibição expressa à raspagem de dados e uso de IA.²⁰¹ A Elsevier, responsável pela base SCOPUS, impõe termos similares.²⁰² Note-se que ambas as empresas possuem produtos baseados em inteligência artificial que utilizam os conteúdos disponibilizados nas referidas plataformas.²⁰³²⁰⁴

Quando nos voltamos para a realidade brasileira, a aplicação desses princípios parece encontrar dificuldades similares. É o que sugere a pesquisa de campo empreendida por Guanaes (2020), que consistiu na realização de questionários e entrevistas com editores de periódicos

²⁰⁰ Disponível em: <https://www.webofscience.com/wos/>. Acesso em 27 set. 2024.

²⁰¹ “2. Your Obligations (...)

(e) Restrictions. You must not (i) introduce any malicious software into Clarivate IP or network; (ii) run or install any computer software or hardware on the Products or network; **(iii) scrape data from the Products**; or (iv) disable or bypass any functionality or restrictions within the Products.

(f) Artificial Intelligence. Unless expressly permitted under a mutually agreed Artificial Intelligence Addendum or other written agreement with us, you must not use and access the underlying Clarivate proprietary data from the Products: **(i) with any of your technology platforms or systems, in a manner which includes or involves your application of artificial intelligence**, such as generative artificial intelligence, machine learning, algorithms or language models (“AI Technologies”) (...).” (Clarivate, 2024a. Destaques feitos pelo autor). Note-se que, embora haja menção explícita à IA Generativa, a proibição se dá para qualquer aplicação de IA, complementando a proibição à coleta de dados automatizada via raspagem.

²⁰² “Terms and conditions (...) Using our Services (...)

You may not use Content from the Services in combination with an artificial intelligence tool, (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool).

You may not use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices, or any similar or equivalent manual process, to: (i) continuously and automatically search, scrape, extract, deep link or index any Content (...).” (Elsevier, 2024).

²⁰³ “Artificial intelligence (AI) helps us deliver on this mission.

AI is already an integral part of our business. Our data offers valuable insights that customers would not be able to access without the help of AI models. The data we use is normalized, correlated and translated by AI to provide meaningful information.” (Clarivate, 2024b).

²⁰⁴ “Scopus AI is an AI-driven research tool that uses the Scopus peer-reviewed research repository to help users understand and navigate unfamiliar academic content. Scopus AI generates summaries based on Scopus abstracts with references to help decipher complex content, facilitate deeper exploration, and provide academic insights. Scopus AI is only available as an add-on subscription.” (Scopus, 2024).

científicos na área da saúde e com gestores de repositórios acerca da postura dos mesmos quanto à abertura e compartilhamento de dados de pesquisa ligados a artigos, especialmente no tocante a questões de Direito Autoral. Entre os editores, observam-se algumas contradições: embora a maioria (76,9%) dos respondentes tenha se manifestado favorável à disponibilização pública de dados subjacentes a artigos, apenas 15,4% já adotava políticas nesse sentido, com mais da metade afirmando que o faria no futuro. Também foi observado que, embora a grande maioria (77,8%) dos periódicos envolvidos na pesquisa recomendasse o uso de repositórios específicos para disponibilização de dados, 61,5% não sabia indicar nenhuma base específica. Ademais, havia uma considerável dúvida quanto à possibilidade de indicar bases que aceitassem embargo dos dados – situação na qual o material não poderia ser imediatamente publicado, dando aos autores o privilégio de uso por um tempo determinado. Houve grande divisão acerca da cessão de direitos sobre os dados, com 46,2% dos respondentes defendendo a aplicação de licenças de uso não-exclusivo, uma minoria (cerca de 7,7%) se posicionando pela exclusividade e outros 46,2% se mostrando indiferentes.

Considerando-se que a pesquisa se voltava exclusivamente a periódicos de acesso aberto, estes resultados não parecem muito animadores: embora haja um discurso à primeira vista favorável à ciência aberta, as respostas apontam para a falta de um consenso institucional amadurecido para aplicar esta abertura a questões de gestão de dados de pesquisa. Um dos entrevistados também manifestou uma grande preocupação em relação à apropriação e reuso dos dados por terceiros (Guanaes, 2020, p. 172-183), em consonância com os resultados observados em questionários com pesquisadores (como veremos na seção seguinte).

Entre gestores de repositórios, a pesquisa apontou para uma maioria mais consolidada no que diz respeito ao compromisso com a abertura dos dados de pesquisa e ao papel dos repositórios em promover este movimento, especialmente no que diz respeito a pesquisas financiadas com recursos públicos, e com a obrigatoriedade da divulgação dos dados pelos pesquisadores, embora isso levante ressalvas quanto a questões de crédito da produção científica. A necessidade de uma regulação internacional padronizada de um direito à abertura de dados recebeu respostas fortemente favoráveis dos respondentes, bem como a percepção de um encorajamento para disponibilização dos dados em domínio público (Guanaes, 2020).

Entretanto, a pesquisa, embora tenha se debruçado empiricamente sobre intermediários da circulação de dados de pesquisa, também observa, em seu referencial teórico e entre entrevistados, a necessidade de se atentar também para aqueles que produzem o trabalho científico na origem. Trata-se de uma observação que diz respeito, diretamente, ao trabalho de

universidades e instituições de pesquisa, que figuram como atores cruciais no processo de inovação, e que tocam no seu menor componente: os próprios pesquisadores e de que forma eles percebem seu papel em meio a este processo. É o que abordaremos a seguir.

4.4. A visão do pesquisador

Pesquisadores ocupam um status peculiar enquanto autores contemplados pela legislação de Direito Autoral: via de regra, não auferem sustento com a reprodução ou comercialização de seus trabalhos, já sendo previamente remunerados para produzir – o que os diferencia de um artista visual ou um escritor fora da academia, por exemplo. Ao mesmo tempo, contudo, sua sobrevivência e crescimento profissional estão intimamente relacionados ao reconhecimento de sua produção, gerando uma pressão pela produtividade de cientistas e institutos tradicionalmente manifesta na famigerada expressão *publish or perish* (“publicar ou perecer”, em inglês) (van Dalen e Henkins, 2012).²⁰⁵

Não parece haver unidade entre os próprios cientistas no que diz respeito a como a propriedade intelectual interfere em seu processo de coleta de dados, tampouco a postura dos mesmos em relação ao compartilhamento do material produzido em pesquisa – e, de fato, é possível observar um movimento paradoxal onde a expansão das práticas institucionais de ciência aberta não parece ser integralmente acompanhada pelos pesquisadores. Por exemplo, Savage e Vickers (2009) contataram dez pesquisadores buscando acesso aos dados de pesquisa associados a publicações disponibilizadas em plataformas abertas; apenas um dos contatados atendeu ao pedido de compartilhamento feito pelos autores. Entre as diversas razões apresentadas, figuraram preocupações com autoria e publicações futuras, bem como um desejo “natural” de reter exclusividade sobre um material que levou tempo e trabalho para ser gerado. Também se observou um receio de que pesquisadores futuros utilizem esses dados para prejudicar os trabalhos originais, seja localizando erros ou empregando metodologias alternativas.

Já Andreoli-Versbach e Mueller-Langer (2013) empreenderam uma observação quantitativa das práticas de compartilhamento de 488 pesquisadores no campo da economia. 89,14% não tinham o costume de disponibilizar seus dados. O principal motivo apontado foi uma aparente preocupação em abrir mão de oportunidades para publicar futuros trabalhos com aqueles dados em benefício de concorrentes. O estudo também indica que esta tendência é mais forte

²⁰⁵ O termo teria sido utilizado pela primeira vez por Wilson (1942, p. 63), e se refere à guinada, nos Estados Unidos, de uma cultura acadêmica centrada no reconhecimento da prioridade para uma baseada na produtividade do cientista (Merton, 1957).

entre cientistas mais jovens, com menos estabilidade e recursos – para quem a quantidade de publicações teria mais peso.

Paralelamente, uma série de estudos realizados ao longo da última década por Tenopir e outros pesquisadores buscou empreender uma análise mais profunda, abordando acadêmicos de múltiplas áreas do conhecimento ao longo do tempo. Os resultados mostram que, por volta de 2011, a maioria dos acadêmicos respondentes era relativamente aberta ao acesso a bancos de dados que pertenciam a outros, se facilmente acessíveis (83,3% concordaram com isso, fortemente ou de forma parcial), bem como a tornar alguns de seus próprios dados acessíveis (78%) sem restrições, mas apenas 40,9% compartilhariam todos os dados (Tenopir et al, 2011). Estudos subsequentes da série descobriram que, ao longo da década, essa percepção teve um crescimento constante - mais de 85% estariam dispostos a acessar os dados de outras pessoas e compartilhar seus próprios dados em 2020. Ao mesmo tempo, há uma preocupação constante com o uso indevido de dados ao longo da década observada (Tenopir et al, 2015; 2020). Os estudos também forneceram evidências de que os problemas impostos pela falta de acesso aos dados não afetam igualmente todas as áreas de pesquisa: as ciências sociais e a ciência/engenharia da computação demonstraram um grau maior de preocupação com a questão quando comparadas aos pesquisadores de medicina e biologia (Tenopir et al, 2011).

Nesse mesmo sentido, Wouters e Haak alcançaram conclusões parecidas: em sua pesquisa de campo, 73% dos respondentes afirmaram que se beneficiariam de mais acesso aos dados de terceiros e 69% entendem o compartilhamento como importante em seu campo, mas somente 64% disponibilizariam seus próprios dados e apenas 37% associam o compartilhamento a prestígio ou crédito pelo trabalho. Os autores também constataram que a inclinação para o compartilhamento de dados também varia de acordo com a cultura de uma determinada comunidade científica, e que a abertura de dados não é uma prática corrente em muitos campos. Ademais, os autores apontam que muitas dessas práticas estariam acontecendo principalmente ao nível individual, estando os financiadores e editores significativamente afastados do processo. Ademais, as práticas de organização e documentação são reportadas como sendo um item que requer bastante esforço, com 25% dos respondentes afirmando que foram treinados para tanto e 60% concordando com a ideia de que especialistas em gestão de dados deveriam participar desse processo, e haveria muito mais preocupação em produzir os dados do que em geri-los (Wouters e Haak, 2015, p. 21, 24).

Entretanto, vale destacar que esta não é a única dificuldade apontada pelos pesquisadores no que diz respeito aos direitos sobre os dados: em graus variados, de acordo com a

nacionalidade, a idade e o campo científico, diversos respondentes também mencionaram o ambiente institucional como um fator que afeta a facilidade de compartilhar informações científicas (Tenopir et al, 2011, p. 5; 2015, p. 10-15; 2020, 16).

Esta contradição entre necessidade de abertura e controle sobre o reconhecimento do trabalho e o uso dos respectivos podem ser vistos em múltiplos exemplos: Este último ponto é particularmente exemplificado no editorial de Longo e Drazen (2016), que criticam o que consideram “pesquisadores parasitas”, que estariam “roubando a produtividade” de outros acadêmicos e usando os dados dos estudos originais contra eles próprios. Lógica similar motiva o surgimento de plataformas como a GIS AID (mencionada no Capítulo 1), que manifesta uma visão mais crítica de modelos de acesso aberto que não conferem salvaguardas e garantias aos interesses dos autores das pesquisas.²⁰⁶

Cabe notar, contudo, que os estudos observados se dão em um contexto de caráter mais global, sem um enfoque específico para pesquisadores brasileiros de forma mais qualitativa. Além disto, entendemos que, embora o enfoque quantitativo das análises acima facilite resultados mais generalizáveis, uma abordagem de cunho qualitativo seria benéfica para complementar estas informações e entender melhor os mecanismos institucionais e comportamentais por trás dos resultados discutidos acima. Desta maneira, procedeu-se a uma pesquisa de campo adicional, com o intuito de verificar de que maneira a experiência brasileira manifesta as questões de acesso, gestão e titularidade sobre dados de pesquisa observadas na literatura aqui mobilizada.

4.5. Acesso a dados de pesquisa: a visão de profissionais brasileiros

Conforme descrito na introdução deste trabalho, os resultados discutidos a seguir são resultado de uma coletânea de entrevistas em profundidade realizadas no contexto de um trabalho mais amplo acerca do uso de dados no ambiente de pesquisa, realizado com diversos países da América Latina entre agosto e novembro de 2021 (Valente, Lana, A. e Houang, 2024, no prelo). Embora o estudo original aborde profissionais de diversos países, este trabalho se

²⁰⁶ “Public Domain archives, where access and use of data takes place anonymously, neither offered any protection of the owners' intellectual property rights to the data, or any other valuable incentive to incentivize the sharing of data, such as transparency on the use of data or effective mechanisms that would ensure the acknowledgement of the owners of the virus and recognition of the submitters of the data. Another hurdle responsible for the lack of rapidly sharing influenza data was in part due to the concern of scientists that they were often not credited for their contribution of data, and to no lesser extent, their worry of being ‘scooped’ by those publishing a manuscript first, without their consent.” (GIS AID, 2024a).

debruça mais especificamente sobre aqueles situados no ou oriundos do Brasil – em consonância com o escopo da presente pesquisa, voltada para o âmbito nacional.

Mais especificamente, esta análise visou extrair mais detalhes acerca de pontos levantados pelos resultados das pesquisas levantadas ao longo deste capítulo, observando a presença de convergências e/ou divergências, além de explorar possíveis informações adicionais.

De um total de 31 (trinta e um) profissionais contatados na área de pesquisa com dados, 9 (nove) responderam aos pedidos de entrevista, dividindo-se entre as seguintes ocupações/setores:

QUADRO 6: Ocupações das pessoas entrevistadas selecionadas para o trabalho

	Pesquisador (a)	Instituição de pesquisa	Agência de fomento	Sociedade civil	Entidade de preservação	Jornalista
Pessoa Entrevistada 1	X					X
Pessoa Entrevistada 2	X					
Pessoa Entrevistada 3				X		
Pessoa Entrevistada 4					X	
Pessoa Entrevistada 5	X					
Pessoa Entrevistada 6					X	
Pessoa Entrevistada 7						
Pessoa Entrevistada 8			X			
Pessoa Entrevistada 9		X				

Fonte: Elaborado pelo autor

Mais propriamente, faz-se uso de técnicas de análise de conteúdo (para mais detalhes, consultar a seção de metodologia na introdução deste trabalho), visando localizar temáticas principais emergentes nas falas de diferentes profissionais dentro do ambiente de pesquisa

brasileiro, com o objetivo de destacar questões centrais a serem consideradas na discussão de uma regulação para o acesso e uso de dados neste tipo de atividade. Foram selecionados profissionais em posição de representantes de equipes de pesquisa em universidades e instituições que lidam corriqueiramente com pesquisa científica intensiva em dados ou o seu fomento.

Desta forma, procedemos à construção das categorias de análise e à codificação de elementos das entrevistas. Os resultados deste processo serão mostrados a seguir, de acordo com a seguinte estrutura: primeiramente, explica-se a categoria ou subcategoria de análise; em seguida, mostram-se exemplos nas falas das pessoas entrevistadas e, por fim, discutem-se os resultados, seção por seção. Após a exposição das categorias principais, levantamos questões adicionais suscitadas por algumas das pessoas entrevistadas, as quais não se fizeram presentes em todas as respostas, mas que guardam relação com os temas selecionados e podem servir de referência para futuras explorações empíricas.

Uso de dados: esta categoria se refere a como os entrevistados incorporam o uso de dados em suas atividades, especialmente de maneira mediada pela informática. Aqui consideramos não apenas a análise de dados propriamente dita, mas também outras práticas que lidem com a utilização de dados ou repositórios no ambiente digital.

QUADRO 7: Exemplos de respostas dentro da categoria “uso de dados”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Então, o jornalismo de dados acaba saindo dessa premissa de você sempre trabalhar com uma grande base de dados, mas necessariamente não precisam ser esses dados estruturados abertos, né?”
Pessoa Entrevistada 2	“(…) a minha linha [de pesquisa] é informação e tecnologia, e eu tenho trabalhado com arquitetura da informação e arquitetura de dados, né?”
Pessoa Entrevistada 3	“E daí basicamente a gente ou produz bases de dados a partir de raspagem de dados públicos ou coleta informação, né?”
Pessoa Entrevistada 4	“A gente está falando de coisas analógicas, que precisam ser digitalizadas e informatizadas para poderem ser disponibilizadas e acessadas e a gente tem uma imensa dificuldade nisso.”
Pessoa Entrevistada 5	“Nós estamos usando os dados das editoras, que a gente pede permissão para trazer todas as referências bibliográficas desses trabalhos publicados (...).”
Pessoa Entrevistada 6	“O que a gente está implementando [na instituição de conservação] é um processo de conservação preventiva, que muitas vezes é associado a um processo de restauração dos bens materiais, então a gente vai digitalizar um acervo.”
Pessoa Entrevistada 7	“Nesse momento, a gente está atuando tanto a na parte de padronização de alguns procedimentos relacionados ao povoamento do repositório institucional da [instituição de pesquisa] e uma outra frente que a gente está iniciando, que é a frente que eu trabalho um pouco mais, que a questão dos dados de pesquisa científica na [instituição de pesquisa].”
Pessoa Entrevistada 8	“Quando é uma pesquisa, um pesquisador geralmente pede antes: ah eu tenho que baixar tudo que existe... eu fiz um robô para baixar tudo que existe nesse tema aí ele entra em contato antes e avisa.”
Pessoa Entrevistada 9	“Então a gente tem que atuar muito próximo dentro dessa nova realidade, que a gente tá falando de dados, né? De mineração de textos, né? Então, assim, é uma área de pesquisa, mas que envolve toda uma discussão aí de dados, de informação e de comunicação. Então, assim, a gente tem muito contato..”

Fonte: elaborado pelo autor

Nota-se que a maior parte dos respondentes manifestou comentários nesse sentido em dois pontos: no início, quando perguntados sobre sua área de atuação, e quando indagados a respeito do uso de ferramentas automatizadas, como inteligência artificial, no seu ambiente de trabalho.

Aqui, também se observa que raramente houve menção expressa a ferramentas de IA – embora haja casos, como expressos pela Pessoa Entrevistada 8, do uso de “robôs”. Entretanto, as respostas sugerem uma multiplicidade de práticas corriqueiras no âmbito da pesquisa que envolvem atividades relacionadas à coleta e análise de dados, tais como digitalização de materiais, raspagem automatizada e montagem e organização de bases – reconhecidas mesmo entre os que não mencionam tais práticas de forma explícita (ex.: Pessoa Entrevistada 5).

Acesso: esta categoria diz respeito a questões relacionadas à localização e obtenção de materiais de pesquisa, incluindo dados. Trata-se de um termo de forte menção entre os entrevistados: o número mínimo de vezes em que a palavra surgia em uma mesma entrevista é de 4 (quatro) vezes, mas todas as demais revelam ao menos 10 (dez) aparições, com alguns entrevistados chegando a mencionar esta palavra cerca de 60 (sessenta) vezes. Por esta razão, optou-se por dar-lhe uma categoria de análise própria.

QUADRO 8: Exemplos de respostas dentro da categoria “acesso”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Mas, ao mesmo tempo, eu vejo essas lógicas predatórias de que você precisa pagar para ter acesso ao artigo, né. Ou você precisa estar ligado à uma, a um instituto de pesquisa, por exemplo.”
Pessoa Entrevistada 2	“Então, se, por exemplo, hoje eu não tivesse o vínculo com a pós-graduação, eu não teria acesso à algumas bases de dados. E se eu não fosse funcionári(a) de uma universidade pública, eu teria ainda menos acesso, né?”
Pessoa Entrevistada 3	“Eu já vi negativas de pedidos de acesso à informação. Fundamentada em direitos autorais, mas para relatórios, geralmente relatórios quando é com consultoria externa, mesmo que seja uma consultoria doada pro poder público, às vezes os relatórios, os produtos estão protegidos.”
Pessoa Entrevistada 4	“Então eles vão dizer para você que você não pode fazer nada sem ter uma expressa autorização da família e de todos os envolvidos. E isso absolutamente não funciona para instituições que têm acervos massivos e tem por obrigação colocar isso para fora e dar publicidade para esses acervos.”
Pessoa Entrevistada 5	“É um dado que está aberto, mas que a gente tem que ter eles, receber da forma que a editora possa fornecer para nós, para a gente ter condição de usar.”
Pessoa Entrevistada 6	“Esses acervos estão cada vez menos digitalizados, cada vez menos acessíveis. Existe um esquecimento, um projeto de esquecimento coletivo que usa o direito como base.”
Pessoa Entrevistada 7	“Essa é a nossa principal fonte de informação, que a gente tem esse acesso garantido por ser uma universidade federal pública.”
Pessoa Entrevistada 8	“(O)s governos e agências e instituições de pesquisa estão exigindo que os pesquisadores publiquem em acesso aberto e os pesquisadores dizem o seguinte: a gente só consegue publicar em acesso aberto se agências de fomento nos financiarem.”
Pessoa Entrevistada 9	“Porque a gente conseguiu aqueles dados, a gente digitalizou, mas a gente queria deixar disponível, porque a gente entende que aquele acervo é super rico, mas a gente tem, a gente entendeu que a gente não tinha autorização para aquilo.”

Fonte: elaborado pelo autor

Aqui, nota-se que, com exceção da Pessoa Entrevistada 5 (que, de forma geral, lida com materiais próprios ou de fácil acesso), os respondentes relatam, de forma geral, empecilhos para o acesso e uso de materiais nas atividades de pesquisa e conservação em função de questões de propriedade intelectual, especificamente de Direito Autoral.

Uma questão recorrente é a necessidade de ligação institucional para que esse acesso ocorra. Diversas das pessoas entrevistadas afirmaram que, se não estivessem afiliadas a uma

universidade, seria muito difícil ou mesmo inviável que pudessem conduzir seu trabalho de pesquisa.

Neste ponto, alguns dos respondentes chamam atenção ao fato de que por vezes, recorre-se a meios ilícitos para a obtenção desse acesso. Uma das pessoas entrevistadas, embora tenha declarado se ater aos recursos aos quais possui acesso institucional, destacou que o recurso à pirataria não seria algo isolado, mas recorrente entre pesquisadores como alternativa para a falta de acesso:

“(…) eu acho que é isso, inclusive, que aumenta o interesse das pessoas por pirataria, não é? Porque as pessoas, elas buscam formas alternativas para encontrar gratuitamente o que elas não encontram de maneira ‘legal’. Vamos dizer assim, né?” (Pessoa Entrevistada 2)

Adicionalmente, observa-se uma questão preocupante no que diz respeito à disponibilização a materiais produzidos pelo (ou para) o Poder Público: o relato da Pessoa Entrevistada 3 sugere uma subsunção dos interesses estatais ao instituto do Direito Autoral, independentemente do interesse público em relação à publicidade desses materiais. Também se nota um movimento similar nos acervos de entidades de conservação, como relatam as Pessoas Entrevistadas 4 e 6 – ambas apontando para um ambiente regulatório que traz efeitos deletérios para a concretização do dever institucional desses repositórios e a manutenção do patrimônio cultural.

Nesse sentido, percebe-se não apenas uma falta de abertura no acesso a dados públicos – algo reiterado por outros respondentes, como a Pessoa Entrevistada 1 – como também um reflexo do processo de tragédia dos *anticommons* materializado na privatização do conhecimento (Argyres e Liebeskind, 1998; Nelson, 2003; Orsi e Coriat, 2006; Ostrom e Hess, 2007b; Hess, 2008; Coriat, 2013).

Autoria e titularidade: nesta categoria, observaram-se asserções acerca da percepção de haver (ou dever haver) alguma forma de autoria, ou qualquer tipo de titularidade, sobre dados de pesquisa e os conjuntos e repositórios onde estão contidos. Trata-se de um tópico que foi perguntado diretamente aos entrevistados, como forma de observar não apenas o grau de entendimento acerca da proteção concedida por lei a esses materiais, mas a visão das pessoas entrevistadas a respeito de se (e como) o Direito deveria proteger este tipo de conteúdo.

QUADRO 9: Exemplos de respostas dentro da categoria “autoria e titularidade”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Eu gosto muito dessas redes informais de pesquisa em que a gente tem a oportunidade de quando a gente pega um material de pesquisa, informar o autor para o que que você vai usar aquilo, né?”
Pessoa Entrevistada 2	“Isso que é interpretação, e isso que é contextualização é talvez um processo intelectual mais presente, mais concreto. E em havendo um trabalho intelectual nisso, eu entendo que deveria haver condição de... não sei se de autoria, mas de alguns tipos de reconhecimento, né?”
Pessoa Entrevistada 3	“No caso de poder público, a gente defende que não tenha, que exista sim uma declaração explícita de licença livre, que isso dá uma segurança jurídica para quem vai usar? Mas que questão de autoria dessas bases de dados, um trabalho criativo, um trabalho original, aí eu já acho que isso colocaria mais barreiras e mais conflitos do que benefícios, né?”
Pessoa Entrevistada 4	“(A)s instituições se sentem, sim, donas únicas donas e, na verdade, o que é mais complicado e que está um pouco por trás desse problema não só como o único dono, mas como o único capaz de interpretar ou de oferecer uma interpretação adequada para isso, né?”
Pessoa Entrevistada 5	Eu acho que todo pesquisador se orgulha de dizer: eu fui o primeiro a fazer, eu fui o primeiro a descobrir.
Pessoa Entrevistada 6	Elas servem à produção de conhecimento. Logo somos coautores da produção de conhecimento.
Pessoa Entrevistada 7	“O que a gente lida dessas pessoas: ‘ai eu estou protegendo meu trabalho’.”
Pessoa Entrevistada 8	“O Direito Autoral ele traz maior segurança para o pesquisador, então ele se sente mais seguro em compartilhar porque ele sabe que ele está protegido (...).”
Pessoa Entrevistada 9	“Eu acho que a atribuição, a atribuição da autoria é do pesquisador, mas que a gente fica tendendo a considerar que isso, que a propriedade é institucional.”

Fonte: elaborado pelo autor

Mais especificamente, quando perguntadas a respeito da proteção e gestão dos dados de pesquisa, notou-se, entre a maioria das entrevistadas, um sentimento geral de propriedade sobre o material produzido. Nota-se que esta percepção nem sempre foi associada a uma noção de originalidade, emanando frequentemente do entendimento de que o trabalho intelectual e técnico envolvido na produção, organização e curadoria dos dados justificaria a garantia de algum controle sobre os dados, como forma de trazer segurança a quem conduziu este trabalho.

Ainda nesta categoria, outro ponto que chama atenção é que, embora pareça haver um predomínio no sentido de considerar como autor aquele que foi responsável pela criação dos dados e/ou a construção da base de dados, observou-se uma certa variedade de possíveis autores e titulares – os quais, nota-se, tendem a reclamar para si titularidade sobre alguma parte do processo de pesquisa. Pesquisadores se sentem donos das suas coletâneas, independentemente de haverem gerado os dados; instituições de pesquisa reclamam a responsabilidade sobre questões patrimoniais e de licenciamento em função do vínculo trabalhista ou apoio financeiro e

organizacional fornecido; museus reclamam controle sobre os acervos em razão do trabalho envolvido na coleta e curadoria do material. Observa-se, então, um possível conflito de interesses, que pode se refletir em diferentes arranjos contratuais ou na fragmentação de direitos (ex.: “a autoria é do pesquisador, mas a propriedade é da instituição”).

À primeira vista, esta fragmentação poderia ser entendida como uma aplicação do paradigma do “feixe de direitos”, mas não implica, por si só, em um modelo de gestão eficiente. Afinal, para Ostrom (1990), o que importa não é a distribuição dos direitos, mas *como* ela é feita. Assim, o uso inteligente desse sistema requer um esforço coordenado e transparente entre os beneficiários, preferencialmente facilitado pelo apoio do Estado. Uma situação de aparente desacordo como a vista acima parece contribuir menos em benefício dos envolvidos e mais como um agravante para as questões de acesso e uso vistas acima, bem como para o cenário de incerteza regulatória questionado mais adiante.

Compartilhamento: se a categoria de acesso, por um lado, expõe diversos entraves enfrentados pelos pesquisadores no que diz respeito ao uso de dados para fins de pesquisa e preservação, a existência de uma noção de titularidade também nos leva a uma questão subjacente: a abertura de pesquisadores e outros profissionais em relação ao compartilhamento desses dados.

QUADRO 10: Exemplos de respostas dentro da categoria “compartilhamento”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Então tem sempre o cuidado de disponibilizar os dados que realmente estejam abertos ou que realmente estejam aí consolidados de alguma forma.”
Pessoa Entrevistada 2	“Porque o discurso que eu escuto é assim: ‘as pessoas não querem abrir os seus dados’.”
Pessoa Entrevistada 3	“Dentro da ciência aberta, que é o compartilhamento das bases de dados, que está ficando cada vez... É uma discussão que está se ampliando, que tem diretrizes, é, mas que ainda precisa evoluir bastante, né?”
Pessoa Entrevistada 4	“Então, quanto mais acessível, mais compartilhado tiver e menos fechado na instituição, mais seguro está aquilo.”
Pessoa Entrevistada 5	“(S)e eu publico um trabalho dizendo que é inédito e que ele é vai contribuir para a ciência, mas um outro grupo que quiser estudar, complementar respostas que eu não dei, acrescentar coisas que eu não fiz, mas não tiver acesso a esses dados, não puder, pelo meu trabalho, reproduzir, eu não acrescentei nada na literatura científica.”
Pessoa Entrevistada 6	“Ao invés do paradoxo entre proteção e acesso, que assim que ele era chamado, sobretudo até o século 21, a gente está entrando num novo paradigma de proteção e acesso, que é justamente quanto mais acessível o objeto é, mais protegido ele está.”
Pessoa Entrevistada 7	“Não posso dizer para você que: ‘ah, não, há um consenso de que os trabalhos devem ficar aberto.’ Tem pessoas dentro da discussão que acham que se a pessoa não quer publicar, é um direito dela também de falar: ‘não, eu não quero publicar nunca’.”
Pessoa Entrevistada 8	“[Os pesquisadores], geralmente, eles são relutantes em compartilhar.”
Pessoa Entrevistada 9	“A gente está entendendo que todos os parceiros, não exatamente pessoa, mas instituições, pelo menos ou líderes dessas pesquisas, precisam estar informados também desse compartilhamento.”

Fonte: elaborado pelo autor

Aqui, nota-se que houve diversas respostas no sentido de que dados, ao menos no âmbito da pesquisa científica, devem ter seu compartilhamento incentivado enquanto prática padrão. Dois dos respondentes, oriundos de instituições de preservação, trouxeram como justificativa um argumento que coloca os acervos de museus, bibliotecas e arquivos enquanto *knowledge commons*, seguindo a lógica apontada por Coriat (2012) de que a proteção do conhecimento está diretamente associada ao seu compartilhamento.

No entanto, uma resposta comum entre os respondentes, mesmo entre a maioria dos que se posicionam de forma favorável ao compartilhamento como princípio científico, é a de que há uma relutância, por parte de muitos pesquisadores, em fazer esse movimento de abertura – e que o fato de serem autores dos trabalhos acadêmicos lhes daria legitimidade para restringir o compartilhamento.

Explorando as respostas mais a fundo, as justificativas para essa resistência são variadas. A principal delas envolve a preocupação em garantir a apropriação sobre os frutos do

trabalho empregado na coleta, produção e armazenamento dos dados – objeto materializado em todas as entrevistas e abordado de forma mais específica na subcategoria “Apropriação”. Dois outros motivos se destacam, contudo. O primeiro deles é a preocupação com a integridade dos dados, sua consolidação e o seu possível desvirtuamento, tópico particularmente destacado pelas Pessoas Entrevistadas 1 e 10:

Eu acho que tem que ter esse momento de curadoria mesmo, né? Do que todos seus dados que podem ser disponibilizados dessa forma e daqueles que precisam de um certo tipo de tratamento para que eles não sejam usados aí de uma forma mal-intencionada ou que possam até invalidar própria pesquisa, né? (Pessoa Entrevistada 1)

(...) muitas vezes, principalmente na área da saúde, a abertura nem sempre é a principal questão, mas essa questão desse compartilhamento. Um compartilhamento em que de um lado garante que vai fazer uso daquela base para pesquisa, mas vai se responsabilizar pelo também, pelo mau uso daquilo que pode acontecer e a gente sabe quanto isso é prejudicial para outros tipos de direitos, né? (Pessoa Entrevistada 10)

A segunda questão trazida por outras pessoas entrevistadas é de ordem mais prática: o próprio uso de um sistema de proteção como mecanismo de manutenção técnica de repositórios. Trata-se de uma preocupação que não aparece necessariamente atrelada à proteção da propriedade intelectual em si, mas a assuntos financeiros de manutenção das bases de dados. Dois respondentes destacaram este ponto:

O argumento geralmente, dos governos é: existe um custo para produzir essa infraestrutura, para entregar o dado como um serviço. Eu não estou cobrando pelo dado, mas eu preciso cobrar pelo serviço, porque essa minha infra-estrutura de manter uma API, o custo de servidor toda vez que você se conecta com uma API para consumir dados, você... E faz uma requisição, você tem um processamento que custa, né? (Pessoa Entrevistada 3)

Então, o que eu penso sobre isso, em sendo mais direto, é que em alguns casos, sim as cobranças são necessárias. É em muitos casos é a única maneira desses objetos existirem. (Pessoa Entrevistada 6)

Nesse sentido, percebe-se um certo paradoxo entre expandir o acesso e, ao mesmo tempo, resguardar direitos sobre o material – o que não passou inteiramente despercebido nas entrevistas, com algumas das pessoas entrevistadas chamando especial atenção para essa tensão:

(...) (E)u defendo que sim, que é o objetivo de serem acessíveis publicamente, gratuitamente é majoritário, é o que nós queremos, mas ao mesmo tempo, também entendo que isso não necessariamente é uma realidade para o momento que vivemos, no lugar em que habitamos. (Pessoa Entrevistada 6)

[O movimento de dados de pesquisa abertos] (é) importantíssimo para o compartilhamento e é também é mais um nicho que se abre para elas [as grandes editoras]. Porque os movimentos de acesso aberto são contraditórios entre si, né. Tem uma questão paradoxal. Tem uns conflitos de interesse. Como isso vai sendo reapropriado, né? Como o capitalismo de plataforma vai ser apropriando disso, né? (Pessoa Entrevistada 8)

No geral, portanto, observa-se que o sentimento de titularidade caminha lado a lado com um interesse predominante de proteger os dados. Entretanto, na maior parte dos casos, esta proteção parece se embasar menos em questões patrimoniais, debruçando-se de forma mais

destacada sobre problemas de atribuição e de preservação dos dados contra usos entendidos como inapropriados, que venham a produzir informações incorretas ou mesmo danosas – as quais podem até mesmo desabonar os próprios pesquisadores que produziram os dados originalmente. São considerações que corroboram as peculiaridades do mercado de trabalho acadêmico e da produção científica descritas anteriormente, constituindo fatores que necessitam ser considerados no que diz respeito à regulação de propriedade intelectual envolvendo dados de pesquisa.

Dados públicos: Esta subcategoria aborda a percepção dos respondentes no que diz respeito, especificamente, a como se deve dar o compartilhamento de bases e de dados de pesquisa quando seu processo de criação envolve investimento do Poder Público.

QUADRO 11: Exemplos de respostas dentro da categoria “dados públicos”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“(N)a lei de acesso a informação tem esse pressuposto de que a informação, o dado aberto no Brasil tem que ser a regra e não a exceção, né?”
Pessoa Entrevistada 2	“Então, ‘nossa, mas eu fiz toda uma coleta aqui, usei um satélite da NASA, aí coletei todos esses dados’, ok? Quem coletou os dados, né? Quem criou os dados foi você?”
Pessoa Entrevistada 3	“Eu defendo sim, tudo que tenha recurso público de pesquisa, por exemplo, acho que tem que estar sujeito também a essas normas de ciência aberta, né? De replicar, de abrir as bases de dados, de retornar isso como um recurso também.”
Pessoa Entrevistada 4	“E isso [a necessidade de autorização] absolutamente não funciona para instituições que têm acervos massivos e tem por obrigação colocar isso para fora e dar publicidade para esses acervos.”
Pessoa Entrevistada 5	“E muito mais do que - não sei quantos por cento - mas deve ser muito mais do que a metade de dinheiro que sustenta a ciência no mundo, vem de verba, vem de dinheiro público. E isso deve retornar para o público.”
Pessoa Entrevistada 6	“(…) a ideia de bens comuns, conhecimento coletivo, de expansão de dados abertas, né? Está ali como se fosse papel de do terceiro setor, papel da organização da sociedade civil organizada e como se isso não estivesse no cerne da questão sobre a constituição das coleções como bem público.”
Pessoa Entrevistada 7	“Considerando as questões éticas que a gente precisa considerar quando trabalha com alguns tipos de dados, eu acho que essas bases são públicas. Então elas deveriam ser protegidas ali, CC0, Creative Commons, porque a gente está fazendo tudo isso com financiamento público.”
Pessoa Entrevistada 8	“Já tem 20 anos que as agências de fomento já têm políticas de acesso aberto mandatórias. Se você recebeu recurso público, você vai ter que publicar em algum lugar pra todo mundo ter acesso aberto.”
Pessoa Entrevistada 9	“Se foi, investimento público, isso precisa de alguma forma manter todo esse percurso”.

Fonte: elaborado pelo autor

Aqui, observa-se uma ressalva pronunciada nas respostas, que destoa do panorama exposto no item anterior: em que pesem os posicionamentos observados em relação à necessidade de atribuição ou algum controle sobre o compartilhamento, todas as pessoas entrevistadas foram enfáticas no sentido de que, se os dados de pesquisa foram coletados ou obtidos com financiamento público (no todo ou em parte), então os mesmos devem ser tornados acessíveis para o público.

As implicações dessa observação se fazem relevantes não apenas quando nos relembramos da presença majoritária de universidades públicas na produção acadêmica nacional, mas também quando se levam em consideração os processos de transferência de tecnologia resultantes de parcerias universidade-empresa, conforme incentivado pela Lei de Inovação. Nesse sentido, colocam-se em destaque as preocupações com a privatização do patrimônio público – conforme colocado pela Pessoa Entrevistada 2, quando esta questiona a legitimidade de se utilizar da autoria de um trabalho para restringir o compartilhamento de material que, ao fim e ao cabo, é oriundo de uma entidade governamental.

Apropriação por terceiros: de forma adjacente à titularidade, observou-se outra categoria relativa ao compartilhamento: as preocupações quanto à forma como o material de pesquisa é utilizado e apropriado por terceiros.

QUADRO 12: Exemplos de respostas dentro da categoria “apropriação por terceiros”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Tem essa questão do reaproveitamento de coisas nossas, né?”
Pessoa Entrevistada 2	“Não, não me lembro agora como que está licenciado, mas está licenciado, e então você sabe que você pode ir lá e fazer esse processo de reuso dos metadados porque tem direito sobre isso.”
Pessoa Entrevistada 3	“Então você vai ter um processamento daqueles dados, vai ter um consumo daqueles dados. Isso custa dinheiro. É legítimo que o governo fique pagando esse serviço para startups ou para empresas?”
Pessoa Entrevistada 4	“E essa coisa que que é, aparece demais, muito mais do que a gente gostaria de ver nas instituições de simplesmente segurar ou sentar em cima dos dados porque você tem que usar eles primeiro.”
Pessoa Entrevistada 5	“Então, a não ser que exista uma proteção por patente, por descoberta, por implementação de uma droga, de um método, de um equipamento, seja lá do que for, os dados devem ser sim, disponibilizados em repositórios de dados e repositórios de trabalhos científicos.”
Pessoa Entrevistada 6	“Quanto mais pesquisa foi feita, maior a condição desse objeto se tornar relevante para Constituição da sua missão última, que é atualizar-se no tempo presente, não é? Levar o passado ao futuro. E isso não é o entendimento majoritário.”
Pessoa Entrevistada 7	“E aí eu lembro que uma das minhas colegas falou assim: ‘Mas e se alguém pegar nosso dados e fazer uma publicação igual?’”
Pessoa Entrevistada 8	“E aí, você tem uma certa resistência que eu acho que acredito ser até justificada, de vários países pobres, periféricos, dizendo: ‘não, a gente não vai publicar nossos dados. Por que a gente vai ficar só fornecendo dados para o norte global?’”
Pessoa Entrevistada 9	“Eu queria explorar melhor isso daqui, porque, por exemplo, se eu deixo isso aqui disponível lá em Harvard ou em Oxford, eles têm 30 estatísticos, não sei quantas pessoas e a gente aqui tem um grupo pequeno e foi muito suado para conseguir isso daqui, esses dados, então a gente queria ter alguma prioridade nessa exploração.”

Fonte: elaborado pelo autor

Nesse sentido, observa-se uma variedade de respostas. Algumas das pessoas entrevistadas manifestaram preocupação com possíveis instâncias de “roubo de dados” decorrentes de processos de abertura, especialmente por parte de instituições ou pesquisadores estrangeiros. Desta forma, a defesa dessas restrições normalmente gira em torno de evitar o plágio ou de garantir a primazia na produção acadêmica.

Desta forma, as respostas das entrevistas demonstram um zelo, por parte de muitos pesquisadores brasileiros, em limitar o compartilhamento de seus dados como forma de garantir que os frutos de seu trabalho sejam devidamente revertidos na forma de publicações e reconhecimento para aquele pesquisador – fatores tidos como essenciais para a manutenção e crescimento na carreira acadêmica –, impedindo que terceiros cresçam à revelia dos envolvidos na coleta e tratamento dos dados. Observa-se, portanto, uma correspondência entre as respostas

nesta seção e os resultados dos levantamentos de Savage e Vickers (2009), Tenopir et al (2011, 2015, 2020) e Andreoli-Versbach e Mueller-Langer (2013).

As Pessoas Entrevistadas 9 e 10, inclusive, trouxeram uma questão adicional relativa ao próprio modelo internacional vigente de dados abertos na ciência: com a adesão de grandes editoras à ciência aberta e o número de periódicos que demanda dos autores o compartilhamento dos dados de pesquisa, os respondentes observaram que, em que pese os efeitos positivos para transparência e avanço em pesquisas, haveria também um temor, por parte de pesquisadores do Sul Global, de que essa abertura e compartilhamento em bases internacionais acabe favorecendo uma relação neocolonial, na qual países periféricos assumiriam a posição de fornecedores de dados a serem simplesmente apropriados e explorados por plataformas estrangeiras, sem contrapartida:

(...) (O)s pesquisadores hoje eles estão falando: “nós somos a favor do acesso aberto, a gente quer que [a agência de fomento] pague APC. Nós vamos pagar APC para Elsevier, para a Springer, para essas editoras, que já são multi-bilionárias. Então, ao invés da gente fortalecer Scielo, La Referência, todos os periódicos universitários das universidades públicas que eles nasceram públicos aqui no Brasil, na América Latina. A gente vai rebaixar isso aqui para dar recurso, ainda mais, para os periódicos privados. (Pessoa Entrevistada 8)

Não tem dinheiro suficiente para a gente ficar bancando a APC no nível que está, no valor que está das APCs né? Para ter um fator de impacto interessante, para um tipo de pesquisa. E aí a gente hoje, né, vê que teve todo um processo aí de apropriação do movimento do acesso aberto. (Pessoa Entrevistada 9)

Tratam-se de colocações que, em outras palavras, expressam uma preocupação não apenas com a apropriação do próprio conteúdo, mas da captura da ciência aberta si por grandes titulares, facilitada pela permanência de desigualdades em relação ao prestígio científico e à medição da produção científica pautada no mesmo – conforme comentado anteriormente. Os relatos acima, portanto, sugerem a incidência de complicações que vão mesmo para além da simples atribuição de direitos exclusivos, trazendo à tona questões contratuais e dinâmicas políticas que exigem uma visão regulatória mais ampla.

Clareza regulatória: por fim, outro elemento notável entre as entrevistas foram considerações acerca do grau de clareza garantido pela legislação de propriedade intelectual àqueles que pretendem trabalhar com dados. Aqui, observa-se o retorno de uma categoria de análise já observada em relação a práticas de mineração de textos e dados no âmbito das empresas brasileiras (Alvarenga, 2019). Desta vez, no entanto, voltamo-nos para a percepção acerca do grau de compreensão que se tem em relação aos limites estabelecidos pelo Direito Autoral no que diz respeito à proteção de dados e o uso dos mesmos para fins de pesquisa.

QUADRO 13: Exemplos de respostas dentro da categoria “clareza regulatória”

Respondente	Exemplo – unidade de contexto
Pessoa Entrevistada 1	“Mexe um pouco com a propriedade intelectual no sentido de que acho que deveriam ter parâmetros oficiais de como você pode ter acesso a esses dados.”
Pessoa Entrevistada 2	“Então, isso é claro assim, no convívio com outras pessoas que fazem pesquisa e o que parece é que as pessoas têm aquela falsa percepção de que o que está disponível aqui, eu não preciso colocar o meu cartão de crédito, é acesso aberto, né?”
Pessoa Entrevistada 3	“E geral, o problema é que se... A gente não costuma tratar na chave de Direito Autoral, mas tem uma falta de clareza na regulação, na legislação sobre a declaração de licenças dessas bases de dados.”
Pessoa Entrevistada 4	“Nessa minha trajetória que eu diria que assim é uma mistura de coisas do tipo: “não conhecemos o suficiente para falar alguma coisa”, então tem uma certa insegurança, né?”
Pessoa Entrevistada 5	“(G)eralmente essas editoras que solicitam elas não publicam se você não disponibilizar os dados. Mas, também, elas são muito claras em dizer: desde que haja solicitação e explicação do porquê esses dados vão ser usados, ou para que esses dados vão ser usados.”
Pessoa Entrevistada 6	“A gente tem que ir atrás mesmo, tem que aprender na prática, muitas vezes aprende rápido, aprender em partes porque isso não está consolidado ainda, nem legislação.”
Pessoa Entrevistada 7	“Ela tem essa insegurança de: ‘ai, eu posso usar esses dados? Porque eles estão protegidos por direitos autorais ou por qualquer outro tipo de proteção.’”
Pessoa Entrevistada 8	“E aí se vai ter Direito Autoral, se não vai, como que vai ser? Você vai usar um conjunto de dados e vai referenciar o autor daquele dado? Ou aquele dado é veio de entrevistas, veio de coisa de laboratório, do rato, isso está muito nebuloso ainda para gente, pra mim.”
Pessoa Entrevistada 9	“Essa não clareza se pode, se não pode, sempre é uma barreira para o uso de bases que a gente acredita que possa ter ali incidindo direitos autorais.”

Fonte: elaborado pelo autor

Um consenso aparentemente geral entre as pessoas entrevistadas é a falta de clareza regulatória acerca do que é permitido em relação a dados de pesquisa e de como lidar com a gestão dos mesmos. A principal manifestação deste problema se traduz na incerteza frequente em relação a quais conteúdos têm seu uso restrito por Direito Autoral e quais os parâmetros de abertura dos dados – mesmo em repositórios públicos, como sugere a Pessoa Entrevistada 1.

Nesse sentido, a Pessoa Entrevistada 5 indica que, por vezes, este tipo de regulação da disponibilização se dá na esfera privada, por meio de disposições contratuais. Por um lado, isto sugere algum nível de segurança na ausência de um posicionamento estatal; por outro lado, isto inspira certa preocupação quando se tem em mente as práticas de grandes editoras em relação à apropriação dos dados de pesquisa a elas cedidos, bem como seus termos de uso pouco “convidativos” no que diz respeito ao uso desses dados para processos de mineração de textos e dados.

Todo este cenário também conta com certos fatores que trazem complicações adicionais: para além dos problemas de atribuição discutidos na categoria de titularidade, também se observa que, embora certas instituições possuam políticas de gestão dos seus dados (a exemplo da Fiocruz, conforme visto acima), algumas das entrevistas (como se vê nas Pessoas Entrevistadas 4 e 8) indicam uma certa deficiência, por parte de outras entidades, de discussões mais aprofundadas envolvendo gestão de dados. A falta desta clareza legal pode levar a processos constantes de tentativa e erro (como indica a Pessoa Entrevistada 6) ou mesmo a uma postura mais defensiva, proibitiva por padrão:

Por exemplo, quando você contrata uma consultoria jurídica, você está esperando que esse dia advogado ofereça as condições para que a instituição esteja protegida e aí o foco é sempre isso, na proteção e ele descamba quase sempre pra uma coisa que é conservadora. Porque o que você falou, na dúvida, a gente não faz e a nossa legislação deixa várias dúvidas no ar em relação a aplicação de fato do instrumento jurídico. (Pessoa Entrevistada 4)

Já outras entrevistas levantaram um outro elemento complicador: a própria variedade de materiais que podem ser objetos de pesquisa, com questões de tratamento e proteção particulares entre si. A Pessoa Entrevistada 9, como visto acima, exemplifica o problema ao citar diferentes questões envolvendo a simples atribuição de origem do material, com outros respondentes se concentrando sobre a própria natureza e grau de abertura dos materiais em si:

A gente entende que a gente vai ter que olhar com cuidado para cada tipo de pesquisa, porque cada tipo de pesquisa produz ou coleta ou acumula tipos diferentes de materiais. E cada um desses pode ter algum aspecto, por exemplo, de sigilo ou de abertura diferente um do outro (Pessoa Entrevistada 9)

Assim, as respostas acima reforçam a problemática já levantada no Capítulo 1 e neste mesmo capítulo, referente não apenas às características que distinguem a produção acadêmica de outros mercados artístico-literários, como também à própria diversidade de práticas e objetos (e portanto, dados) dentro da pesquisa propriamente dita. Trata-se de um desafio

[A lei brasileira] joga tudo no mesmo balaio e os objetos de pesquisa, eles são muito diferentes de outras criações intelectuais, não é? Então acho que precisa conseguir, não sei se dentro da mesma lei ou separar como foi feito para software, talvez conseguir trabalhar as especificidades de cada tipo de criação intelectual, né? E trazer coisas porque simplesmente dizer “existentes ou que se invente no futuro” abre muita possibilidade para tudo e nada, né? (Pessoa Entrevistada 2)

Desta forma, as entrevistas apontam para uma variedade de informações relevantes: em confirmação do que fora observado nos levantamentos quantitativos e mistos discutidos neste capítulo, diversos pesquisadores criticam a dificuldade de acesso a materiais de pesquisa como empecilho à atividade científica. Ao mesmo tempo, porém, há um sentimento de posse sobre os dados e um receio palpável de compartilhá-los com terceiros de forma irrestrita – sob pena de perder o controle sobre o uso correto dos dados, favorecer outros pesquisadores às

custas de si e, no caso de redes internacionais, de perpetuar uma relação neocolonial de fornecimento de “matérias-primas” a países do Norte Global.

Nota-se, contudo, que as preocupações observadas nos pesquisadores tratam, primariamente, da valorização da atribuição e da integridade, com questões econômicas aparecendo de forma secundária. Em grande parte, isso pode ser explicado pelas próprias características do mercado de trabalho acadêmico, no qual o pesquisador típico não recebe valores pela publicação de exemplares (às vezes, *paga-se* para publicar): seu principal sustento não vem, portanto, da exploração econômica das obras, mas do prestígio que deriva da associação do nome do pesquisador ao seu trabalho. Tem-se, portanto, um paradigma muito distinto do tradicionalmente visto na teoria autoralista e nas discussões internacionais, o que demanda uma atenção especial ao se considerar qualquer regulação de propriedade intelectual no âmbito da pesquisa.

4.6. Um ambiente marcado pela complexidade

Este capítulo teve por objetivo analisar o ambiente acadêmico, atentando para práticas referentes à produção científica, seu papel dentro do ecossistema de inovação e de que forma as questões de acesso e compartilhamento de dados surgem neste cenário.

Nesse sentido, observa-se que as universidades e centros de pesquisa são, tanto para a literatura como para o Estado, um personagem essencial no ecossistema de inovação, com especial relevância em contextos econômicos onde tecnologias intensivas em conhecimento ganham um protagonismo cada vez maior. Ao mesmo tempo – e, possivelmente, em consequência desse cenário –, nota-se a prevalência crescente de um modelo de política no qual a missão da academia de criar conhecimento em benefício da sociedade inclui, também, a participação direta da universidade no processo tecnológico e na criação de produtos e serviços. Como apontado por Nelson (2003), foram-se os tempos de uma ciência centrada na serendipidade e no conhecimento por si só: a pesquisa científica, cada vez mais, ganha um foco voltado para a produtividade, a aplicação empresarial e a geração de valor econômico como sinônimos do cumprimento da missão das universidades e instituições de pesquisa em produzir e socializar o conhecimento.

Não surpreende, portanto, a constatação de que a propriedade intelectual parece surgir de forma constante nesse processo, enquanto mecanismo de proteção e tangibilização deste valor percebido no conhecimento, tornando-o um objeto negociável e comercializável. Tampouco espanta que a ciência tenha, com o passar dos anos, se tornado mais um palco para o processo de cerceamento e privatização do conhecimento, diante do qual surgiram os

movimentos de ciência aberta e acesso ao conhecimento, dos quais o acesso a dados constitui um desdobramento mais recente. É uma disputa aparentemente em andamento, com grandes editoras e repositórios demandando um controle cada vez maior sobre os dados de pesquisas que lhes são submetidas, ao mesmo tempo em que repositórios abertos encontram dificuldades no sentido de encontrar um consenso regulatório que permita resguardar os interesses de pesquisadores e, simultaneamente, preservar sua missão institucional de combater a oligopolização do conhecimento.

Ao mesmo tempo, também não podemos nos furtar da constatação de que tanto a literatura como as entrevistas mobilizadas neste trabalho apontam para a percepção da produção intelectual, especialmente entre pesquisadores, como algo carregado de valor pessoal – há uma preocupação com o uso que é feito do material por eles produzido e a prevenção de uma apropriação indevida do mesmo. Trata-se de um dilema persistente no meio acadêmico que parece, em parte, motivado pelas dinâmicas profissional e de mercado na academia: se sua carreira não depende da comercialização de cópias de seu trabalho, mas do seu reconhecimento e exposição, então a pesquisa científica, ainda que não venha a resultar na criação de spin-offs ou contratos com empresas, ainda manifesta, de alguma forma, preocupações de ordem econômica que não são devidamente cobertas pelos direitos patrimoniais padrão encontrados na LDA.

Nesse sentido, observa-se que, mais do que o controle sobre a reprodução, adaptação ou outros direitos exclusivos contemplados diretamente pelo Direito Autoral, o acesso às obras (ou melhor, a dificuldade de obtê-lo ou concedê-lo) aparece como uma questão muito presente entre os entrevistados. Diversos pesquisadores apontaram que a combinação de preços exorbitantes e a falta de opções lícitas para quem não pertence a uma determinada instituição de pesquisa se coloca como um obstáculo à pesquisa. Isso evidencia a permanência de um problema ainda relativo ao movimento de ciência aberta e do acesso a artigos científicos, o que se potencializou com as questões de acesso a dados.

Nesse cenário, a concessão de acesso aos próprios materiais também surge como uma questão mais delicada. O que se observa é uma preocupação com a exploração dos recursos de pesquisa de forma geral, o que inclui repositórios, coleções e bases de dados. Em alguns momentos, observa-se que a democratização da pesquisa é uma pauta geralmente aceita pela maioria dos entrevistados, mas que, quando o uso por terceiros vem à tona, entende-se que esta acessibilidade precisa ser acompanhada por um trabalho de mediação. Isto se dá não apenas por preocupações ligadas ao reconhecimento da autoria e a preservação do material em um nível intelectual, mas também por questões materiais: sejam elas porque o acervo tem valor

econômico para o titular ou porque a manutenção desse acervo, independentemente do interesse público envolvido, implica em custos que precisam ser cobertos de alguma forma.

Ao mesmo tempo, as observações levantadas por algumas das pessoas entrevistadas sobre como a abertura de dados pode se refletir no Sul Global também demandam ponderação. É possível dizer que a abertura irrestrita pode levar à apropriação maior de dados por parte do Norte Global? Nesse sentido, tanto o material bibliográfico quanto as entrevistas e documentos analisados neste trabalho apontam que a adesão à ciência aberta pelas grandes editoras vem acompanhada de diversas ressalvas. Demanda-se dos autores que se forneçam os dados, mas impõe-se um controle considerável sobre como outros autores podem obter esses mesmos dados. Ao mesmo tempo, as preocupações com um processo de colonialismo de dados em curso merecem atenção, especialmente quando se observa que os periódicos com os maiores fatores de impacto são publicados pelo Norte, perpetuando uma dinâmica de apropriação de dados já estabelecida pelas grandes editoras – algumas das quais, como observado, também vêm explorando o uso da inteligência artificial com seus acervos.

Por outro lado, se levarmos em conta tudo o que foi exposto até este momento, também se nota que este cenário parece persistir mesmo diante de uma legislação de Direito Autoral restritiva ao uso de dados, tal como a brasileira. O que se observa, portanto, é um desequilíbrio de poder de barganha potencializado por uma lógica pautada na produção e no fator de impacto – o que nos traz à mesma lógica de conjunção entre ciência e mercado discutida no início. Nesse cenário, indaga-se até que ponto manter uma restrição à mineração de textos e dados no Brasil seria uma solução ou se, na verdade, pouco faria para coibir esse processo de exploração enquanto limitaria a capacidade de o pesquisador brasileiro se colocar como um personagem mais ativo nesse sistema, capaz de acessar e utilizar o material de pesquisa fornecido pelos países mais industrializados.

Por fim, cabe ainda ressaltar duas questões. A primeira, colocada pela literatura, se refere à documentação do material de pesquisa, apontada como um desafio adicional à implementação de um ambiente de produção acadêmica aberta. Embora, à primeira vista, este pareça ser um problema de ordem mais técnica do que regulatória, isto também se intersecciona com o Direito Autoral e provê uma complicação adicional para a mineração de textos e dados na pesquisa científica. Em primeiro lugar, porque a resistência ao compartilhamento de dados, observada nas entrevistas e nos estudos quantitativos, contribui para dificultar a construção e curadoria de repositórios. Em segundo lugar, porque diversas regulações sobre inteligência artificial – implementadas ou em discussão, inclusive no Brasil – preveem a necessidade de

procedimentos de transparência sobre os dados utilizados. Se não há documentação, o cumprimento desse requisito se torna impossível.

A segunda questão, observada tanto na literatura apresentada no Capítulo 1 como os estudos e entrevistas deste capítulo, diz respeito ao fato de que a legislação de Direito Autoral parece inapropriada para lidar com o contexto de produção acadêmica, tão distinto e com uma diversidade tão grande de definições de dado e de objetos de pesquisa – cada um com particularidades também diversas, que possivelmente não cabem no próprio Direito Autoral.

O regulador brasileiro, portanto, se encontra diante de um desafio formidável: garantir um arcabouço jurídico que traga mais holofotes para a prática científica, estabelecendo um regime de propriedade intelectual que entenda o Direito Autoral não como um bem abstrato, mas enquanto política ligada ao incentivo à cultura e ao conhecimento, que estabeleça limites mais claros para a titularidade, acesso e uso de materiais de pesquisa. Resta agora analisar até que ponto as discussões regulatórias no Brasil têm avançado, e até que ponto vêm respondendo aos anseios da comunidade científica.

CAPÍTULO 5: OS DEBATES SOBRE MINERAÇÃO DE TEXTOS E DADOS NO DIREITO AUTORAL BRASILEIRO

Diante dos aspectos técnicos da mineração de textos e dados, das práticas de pesquisa e do panorama internacional observados nos capítulos anteriores, é chegado o momento de se voltar ao ambiente regulatório brasileiro. Mais especificamente, analisado o papel da pesquisa no ambiente de inovação no Capítulo 4, resta observar de que maneira o ordenamento jurídico brasileiro tem se configurado nos últimos anos e, finalmente, de que maneira o ambiente regulatório no Brasil vem recepcionando a mineração de textos e dados, especialmente no âmbito do Direito Autoral. Isto envolve uma análise bibliográfica e, principalmente, documental, que permita observar não apenas a evolução da doutrina e da jurisprudência autoralista no Brasil, como também eventuais desenvolvimentos no campo do Poder Legislativo nesse sentido – o que constitui o foco principal deste capítulo.

5.1. Limitações e exceções no Direito Autoral brasileiro contemporâneo

No geral, observa-se um processo de maior abertura às limitações e exceções como instrumentalização da função social da propriedade – listada no art. 5º da Constituição Federal (Brasil, 1988)²⁰⁷ – e que, à época, já começava a encontrar eco no Direito Autoral em alguma medida. Evidência disso está no julgamento do Recurso Extraordinário n. 113.505-1, realizado em 28 de fevereiro de 1989 pelo Supremo Tribunal Federal (STF), sobre o uso, por uma emissora de televisão, de imagens extraídas de programas de televisão oriundos de um concorrente. Ali, destacou-se a finalidade do uso como elemento determinante para estabelecer o escopo das limitações da proteção autoral, além da importância de o regime de limitações e exceções permitir certa flexibilidade interpretativa, bem como o auxílio da analogia como ferramenta de aplicação do Direito (Brasil, 1989).

Fora do Poder Judiciário, contudo, tal movimento de abertura foi mais tardio. Teóricos como Antônio Chaves – que chegou a categorizar certas L&E em patamar equiparável à censura²⁰⁸ –, Carlos Alberto Bittar – que notadamente defendia exceções taxativas e interpretadas

²⁰⁷ “Art. 5º (...)”

XXII - é garantido o direito de propriedade;

XXIII - a propriedade atenderá a sua função social (...)” (Brasil, 1988).

²⁰⁸ Conforme nos informa a pesquisa documental de Valente (2018, p. 151), o estudioso, em carta enviada durante a tramitação da Lei 9.610/98 no Congresso Nacional, criticava a imposição de “um duvidoso interesse público, num plano inclinado que, se não for sabiamente contido, ameaça não apenas estiolar a justa recompensa que merece a produção intelectual, como ainda dará ensejo a mil e uma formas de locupletamento indevido”, e apontava como restrições ao Direito Autoral a serem removidas não só “a execrada censura”, eliminada com a Constituição de 1988, como “as referências às exigências de natureza científica, didática ou religiosa, às informações de imprensa, às obras de arte existentes em logradouros públicos.”

restritivamente sob o argumento de que “interpretam-se restritivamente os negócios jurídicos sobre os direitos autorais”²⁰⁹–, Eduardo Vieira Manso,²¹⁰ Plínio Cabral²¹¹ e Maria Elaine Rise Jundi²¹² não manifestavam uma posição muito favorável a uma abertura às limitações e exceções, por vezes identificando nelas uma penalização ao titular ou brecha para abusos.

No Legislativo, a situação não era mais branda: Valente (2018), ao discutir o processo de tramitação da Lei nº 9.610 – atual Lei de Direito Autoral brasileira –, fornece um exemplo deste cenário ao apontar para uma versão do projeto posta pelo então deputado José Genoíno durante a tramitação do então Projeto de Lei. A proposta preservava apenas o direito de citação e excluía todas as demais limitações existentes na legislação da época,²¹³ além de trazer menção a um “domínio público remunerado”.²¹⁴ No fim, a versão que viria a ser promulgada não chegou a esse extremo, mas notadamente ampliou o prazo de proteção dos direitos patrimoniais em 10

²⁰⁹ “Por outro lado, (a Lei de Direitos Autorais) enumera taxativamente as hipóteses de uso livre (...). Em todos os casos, prevalece a interpretação estrita, de sorte que sempre devem ser respeitados todos os requisitos expostos, a fim de que tranquila possa ser a sua utilização. A Lei nº 9.610, de 19.02.98, por meio de seu art. 4º, disciplina a matéria dizendo que: ‘interpretam-se restritivamente os negócios jurídicos sobre os direitos autorais’” (BITTAR, 2008, p. 70-72). Cabe notar que diversos usos livres listados no rol de limitações e exceções não constituem negócios jurídicos propriamente ditos (Lewicki, 2007, p. 148-149; 164-165; Souza, C., 2009, p. 181-182), de forma que o referido artigo não parece embasar tal interpretação.

²¹⁰ “(...) nada obstante toda exceção imposta ao Direito Autoral seja um ônus que se exige do autor como verdadeiro castigo, há nessa exclusão um verdadeiro ato de misericórdia, no sentido mais canônico da expressão” (MANSO, 1982, p. 260).

²¹¹ “Elas (as limitações) são específicas e fechadas. Constituem *numerus clausus* e não podem, por isso mesmo, estender-se além daquilo que a lei fixou. Apesar disso, é justamente nessa área que se verificam os maiores conflitos, com a tendência de muitos a aumentar o alcance daquilo que a legislação faculta, gerando-se abusos de toda a natureza” (CABRAL, 2003, p. 70).

²¹² “O objetivo do legislador ao elaborar o texto, tanto da Constituição Federal, quanto da lei que regula a matéria dos direitos autorais, outro não era senão preservar o titular e deixar a critério dele autorizar ou não o uso de sua obra, mediante critérios de conveniência artística, financeira, econômica e moral, resguardando a si a “palavra final” com relação à utilização de sua criação de espírito. Se assim não fosse, os titulares estariam totalmente inertes quanto ao uso de sua criação, ficando a critério do usuário pagar-lhe ou não o valor que entendesse, unilateralmente adequado, retirando do titular a facultas agendi de, inclusive, impedir o uso da obra, em razão de conveniências pessoais” (JUNDI, 2004).

²¹³ A única limitação presente no projeto Genoíno era o direito de citação: “Art. 34. A notícia ou citação de trecho de obra intelectual e a crítica ou polêmica que a tenha por objeto não constituem ofensa aos direitos do criador.” (Valente, 2018)

²¹⁴ “Art. 49. Compete ao Estado, através do Conselho Nacional de Direito Autoral – CNDA, a defesa da obra caída em domínio público.

Art. 50. A exploração econômica de obra de domínio público depende:

I – de prévia comunicação ao Conselho Nacional de Direito Autoral, com especificação de modalidade de uso e número de exemplares, quando for o caso;

II – do recolhimento dos direitos autorais arbitrados pelo Conselho Nacional de Direito Autoral, com base nos valores correntes do mercado” (Brasil *apud* Valente, 2018).

Parágrafo Único – O descumprimento das obrigações previstas neste artigo importa em contrafação.”

anos,²¹⁵ e tornou o rol de limitações e exceções mais restritivo do que antes,²¹⁶ ecoando as transformações maximalistas no sistema internacional de Propriedade Intelectual à época.

Apesar disto, os anos 2000 viram o crescimento de uma doutrina jurídica voltada para o caráter extensivo das limitações, recorrendo justamente à Constituição e ao direito vigente para resgatar o valor do interesse público no Direito Autoral. Pimenta (2004, p. 318), por exemplo, já se posicionava por uma interpretação extensiva não só do rol de usos já permitidos mas para além destes, tomando como base o papel da analogia já previsto na Lei de Introdução às Normas do Direito Brasileiro – e em muito lembrando o entendimento emitido pelo STF em 1989.²¹⁷ Posicionamento similar também seria alcançado por Branco (2007, p. 232), ao entender que tais direitos e princípios legitimariam condutas que não estão necessariamente prescritas em lei – ou, mesmo, que operem de forma aparentemente oposta a ela.

Já A. Souza (2005, p. 286-288), traz para os holofotes a função social da propriedade enquanto expressão manifesta de observância ao interesse público na Constituição, e presente repetidas vezes tanto na lei como na jurisprudência ou em tratados internacionais. Para o autor, as limitações e exceções no Direito Autoral brasileiro são expressão por excelência desta comando constitucional, servindo como garantia da observação de diversos outros direitos fundamentais – principalmente no que diz respeito ao direito à educação, informação e cultura, ao direito do consumidor e a princípios constitucionais mais gerais, como a livre iniciativa e a soberania nacional.²¹⁸

²¹⁵ Aqui, confrontamos o art. 42, §2º da lei nº 5.988/73, que afirma que “Os demais sucessores do autor gozarão dos direitos patrimoniais que este lhes transmitir pelo período de sessenta anos, a contar de 1º de janeiro do ano subsequente ao de seu falecimento” (Brasil, 1973) com o art.42 da lei nº 9.610/98, segundo o qual “Art. 41. Os direitos patrimoniais do autor perduram por setenta anos contados de 1º de janeiro do ano subsequente ao de seu falecimento, obedecida a ordem sucessória da lei civil” (Brasil, 1998b).

²¹⁶ Destacamos, a título de exemplo, a alteração da possibilidade de reproduzir obras no todo ou em parte, no contexto de obra maior, de caráter científico, didático ou religioso do art. 49, I, a (Brasil, 1973), que passou a aceitar a cópia integral somente de obras de artes plásticas, mas que ampliou o uso para quaisquer fins que não comprometam a exploração normal da obra ou prejudiquem injustificadamente os interesses legítimos do autor, e a restrição da cópia privada, que passou a permitir a reprodução somente de pequenos trechos no art. 46, II (Brasil, 1998b).

²¹⁷ “Art. 4º Quando a lei for omissa, o juiz decidirá o caso de acordo com a analogia, os costumes e os princípios gerais de direito” (Brasil, 1942).

²¹⁸ “Fazendo os direitos da coletividade, motivados pelo interesse público, parte da estrutura nuclear destes direitos, não é razoável a aplicação de uma interpretação restritiva aos seus dispositivos, especialmente quando se reconhece a extensividade da interpretação quanto às formas de utilização exclusiva e mesmo aos tipos de obras protegidas. Diante disso, a interpretação mais adequada, e única condizente com os mandamentos constitucionais e por isso juridicamente sustentável, é a extensiva (...). Os interesses sociais da coletividade, constitucionalmente previstos, que refletem na utilização das obras autorais são principalmente de três ordens que consistindo nos direitos à informação, à educação e à cultura. A estes podem ser acrescidos outros cujas relações com os direitos autorais ainda não se encontram suficientemente claras tais como os direitos dos consumidores, principalmente de acesso, à livre concorrência, ao entretenimento ou à soberania nacional, em razão da concentração da titularidade patrimonial em organizações internacionais (...). O interesse público tem uma influência histórica sobre a conformação

Tal posicionamento é contemporâneo ao julgamento do Recurso Extraordinário nº 201.819, julgado em 2005,²¹⁹ sobre a possibilidade de o ECAD, excluir um sócio do seu quadro. Mais interessante é constatar que a lide em questão não tratou de limitações e exceções, mas tornou-se, mesmo assim, um sinal de que o STF daria continuidade à sua defesa da função social da propriedade. Nesse sentido, o Tribunal deixou claro que entes privados não operam à parte dos princípios previstos na lei e tampouco na Constituição – particularmente no que diz respeito a um órgão que exerce função relevante para o interesse público. Em outras palavras, privado e público se revelam como faces da mesma moeda: o atendimento aos direitos fundamentais e aos interesses da coletividade (Brasil, 2005).

Carboni (2006, p. 172; 233), por sua vez, afirma que a interpretação correta do rol de limitações seria a taxativa, mas não só defende a criação de uma cláusula geral que permita a criação de mais usos livres, como destaca a importância das restrições “extrínsecas” ao exercício desses direitos, como a função social da propriedade e dos contratos e o abuso do direito, enquanto elementos que também limitam o Direito de Autor para além do rol de limitações interno à lei autoralista.

Em consonância com a jurisprudência (e em sintonia com as noções pelos autores acima), Lewicki (2007, p. 56-83) se dedica a explicitar a posição do Direito Autoral não como disciplina insular ao restante do Direito, mas como parte desse – estando, portanto, sujeito ao mesmo regime constitucional. Ademais, assim como outros campos do Direito, não está escrito em pedra, mas é, sim, resultado de constantes transformações demandadas pelas mudanças

jurídica dos direitos autorais, fundamentando as suas limitações e derrogações, e demandando equilíbrio e proporcionalidade na conjunção dos diversos interesses que constituem o conteúdo destes direitos (...). A definição da função social dos direitos autorais é o reflexo da expressão dos direitos da coletividade na estrutura jurídica de proteção aos direitos autorais” (Souza, A., 2005, p. 286-288).

²¹⁹ “(...) A ordem jurídico-constitucional brasileira não conferiu a qualquer associação civil a possibilidade de agir à revelia dos princípios inscritos nas leis e, em especial, dos postulados que têm por fundamento direto o próprio texto da Constituição da República, notadamente em tema de proteção às liberdades e garantias fundamentais. *O espaço de autonomia privada garantido pela Constituição às associações não está imune à incidência dos princípios constitucionais que asseguram o respeito aos direitos fundamentais de seus associados.* A autonomia privada, que encontra claras limitações de ordem jurídica, não pode ser exercida em detrimento ou com desrespeito aos direitos e garantias de terceiros, especialmente aqueles positivados em sede constitucional, pois a autonomia da vontade não confere aos particulares, no domínio de sua incidência e atuação, o poder de transgredir ou de ignorar as restrições postas e definidas pela própria Constituição, cuja eficácia e força normativa também se impõem, aos particulares, no âmbito de suas relações privadas, em tema de liberdades fundamentais (...). As associações privadas que exercem função predominante em determinado âmbito econômico e/ou social, mantendo seus associados em relações de dependência econômica e/ou social, integram o que se pode denominar de espaço público, ainda que não-estatal. A União Brasileira de Compositores - UBC, sociedade civil sem fins lucrativos, integra a estrutura do ECAD e, portanto, assume posição privilegiada para determinar a extensão do gozo e fruição dos direitos autorais de seus associados (...). O caráter público da atividade exercida pela sociedade e a dependência do vínculo associativo para o exercício profissional de seus sócios legitimam, no caso concreto, a aplicação direta dos direitos fundamentais concernentes ao devido processo legal, ao contraditório e à ampla defesa (art. 5º, LIV e LV, CF/88). IV. RECURSO EXTRAORDINÁRIO DESPROVIDO.” (Brasil, 2005. Destaque feito pelo autor)

culturais e sociais ocorridas ao longo das épocas. Assim, o atual cenário jurídico e socioeconômico demanda um Direito Autoral menos concentrado em proteger a propriedade dos titulares e mais em garantir o desenvolvimento educacional e cultural.

Já C. Souza (2009, p. 204-277) parte da teoria do abuso de direito existente no Código Civil e aplica ao Direito de Autor: o autor afirma que, nos casos em que a tutela autoralista extrapola os limites impostos justamente por seu propósito econômico, função social, ou mesmo a observação da boa-fé ou dos bons costumes, então o exercício do Direito Autoral, neste caso, deve ser entendido como ilícito, porquanto abusivo, a fim de que o equilíbrio entre os interesses regulados por este domínio do Direito seja observado.

Assim, fica patente um processo de alinhamento entre a doutrina jurídica, progressivamente mais aberta às limitações e exceções como garantidoras da função social da propriedade, e os Tribunais Superiores, defensores da subsunção dos interesses privados ao bem público. Esta visão persistiu em 2011, quando o Superior Tribunal de Justiça julgou o Recurso Especial nº 964.404. No caso em tela, o Escritório Central de Arrecadação e Distribuição (ECAD) efetuou cobrança pela execução de músicas (através de intérpretes e sonorização ambiente) em um evento religioso, sem fins lucrativos, promovido pela Mitra Arquidiocesana de Vitória, no Espírito Santo. Nesta ocasião, o Tribunal foi explícito e enfático ao afirmar o rol de limitações e exceções do Direito Autoral brasileiro como exemplificativo. Desta maneira, estabeleceu-se, efetivamente, uma cláusula aberta para a criação de novas limitações e exceções. Neste caso, o Tribunal estabeleceu um modelo simples, basicamente utilizando a regra dos três passos como cláusula geral para identificação dessas limitações:

(...) II - Necessidade de interpretação sistemática e teleológica do enunciado normativo do art. 46 da Lei n. 9610/98 à luz das limitações estabelecidas pela própria lei especial, assegurando a tutela de direitos fundamentais e princípios constitucionais em colisão com os direitos do autor, como a intimidade, a vida privada, a cultura, a educação e a religião.

III - O âmbito efetivo de proteção do direito à propriedade autoral (art. 5º, XXVII, da CF) surge somente após a consideração das restrições e limitações a ele opostas, devendo ser consideradas, como tais, as resultantes *do rol exemplificativo* extraído dos enunciados dos artigos 46, 47 e 48 da Lei 9.610/98, interpretadas e aplicadas de acordo com os direitos fundamentais.

III - Utilização, como critério para a identificação das restrições e limitações, da regra do teste dos três passos ('three step test'), disciplinada pela Convenção de Berna e pelo Acordo OMC/TRIPS. (Brasil, 2011. Destaque feito pelo autor)

Mais recentemente, a interpretação extensiva das limitações foi consolidada no Enunciado 115 da III Jornada de Direito Comercial, promovida pelo Centro de Estudos Judiciários

da Justiça Federal,²²⁰ deixando uma diretriz clara de conduta para o Judiciário e sedimentando o reconhecimento da função social da propriedade no âmbito do Direito Autoral. Sem dúvida, isto representa um passo importante para que o Poder Judiciário participe na regulação do Direito Autoral nacional por intermédio de demandas diretas da sociedade, permitindo às cortes expandir o rol de limitações e exceções existentes e materializar a função social da propriedade dentro do ordenamento jurídico pátrio.

Contudo, é importante lembrar que, como qualquer cláusula aberta, isto não implica em dizer que o uso de obras na mineração de textos e dados para pesquisa se encontra permitido no Direito brasileiro. Ao contrário, o entendimento do STJ parece reforçar que, para que qualquer uso seja considerado parte efetiva do regime de L&E, ele necessita cumprir a regra dos três passos integralmente – incluindo, como já visto, a criação de um certo caso especial, com condições específicas para que a licitude da prática seja observada. O que o Acórdão e o Enunciado preveem é tão somente que o Judiciário está investido da prerrogativa de definir que caso seria esse, desde que esta demanda lhes seja trazida no caso concreto.

Ademais, observa-se que, ao menos no âmbito do próprio Superior Tribunal de Justiça, este teste ainda não foi utilizado para a definição de uma nova limitação desde a sua criação: uma pesquisa jurisprudencial no site do STJ em 13 de fevereiro de 2024, inserindo “RESP 964.404” como termo de busca, indicou apenas cinco resultados, nenhum dos quais aplica o acórdão no sentido de apontar outros usos livres – os acórdãos encontrados ora são desdobramentos do mesmo processo de 2011 ou tratam, na verdade, da interpretação de limitações e exceções já previstas em lei. Ademais, ainda que algum uso seja considerado permitido em algum tribunal regional, não é seguro presumir que tal decisão tenha efeitos em esfera nacional até que um Tribunal Superior confirme seu conteúdo.

Por outro lado, isto não quer dizer que nenhum avanço ocorreu nas instâncias superiores desde 2011. Ao contrário, ainda é possível observar casos de especial relevância para as limitações e exceções de forma geral, especialmente quando se pensa na extensão destas para a mineração de textos e dados em alguma capacidade. Notadamente, chama-se atenção para o Recurso Especial nº 1450302/RJ, que trata da reprodução, na íntegra, de um poema de Cecília Meireles em um livro didático. No caso em tela, a questão central girou em torno do escopo quantitativo do direito de citação previsto no art. 46 da LDA. Mais especificamente, observou-

²²⁰ “ENUNCIADO 115 – As limitações de direitos autorais estabelecidas nos arts. 46, 47 e 48 da Lei de Direitos Autorais devem ser interpretadas extensivamente, em conformidade com os direitos fundamentais e a função social da propriedade estabelecida no art. 5º, XXIII, da CF/88” (Brasil, 2019a, p. 6).

se se a citação de "passagens de qualquer obra para fins de estudo, crítica ou polêmica, na medida justificada para o fim a atingir" pode contemplar a obra completa. A resposta, perante o Tribunal, foi positiva: o poema em questão ("O Lagarto Medroso") é um texto bastante curto, de forma que, no entendimento dos Ministros, sua menção parcial seria prejudicial à compreensão do texto, podendo levar a interpretações errôneas (e a uma violação do direito moral de integridade do autor).²²¹ Esta interpretação, embora aplicada em um contexto aparentemente específico, tem relevância para atividades de mineração de textos e dados quando lembramos que estas ferramentas não funcionam satisfatoriamente com "pequenos trechos", mas dependem de uma vastidão de materiais para serem capazes de funcionar – quanto mais completos, melhor.

²²¹ "RECURSO ESPECIAL - AÇÃO DE OBRIGAÇÃO DE FAZER C/C PEDIDO CONDENATÓRIO - ALEGAÇÃO DE VIOLAÇÃO DE DIREITO AUTORAL EM RAZÃO DA PUBLICAÇÃO DA INTEGRALIDADE DO POEMA "O LAGARTO MEDROSO" DA ESCRITORA CECÍLIA MEIRELES NO BOJO DE LIVRO DIDÁTICO DESTINADO AO ENSINO FUNDAMENTAL - INSTÂNCIAS ORDINÁRIAS QUE REPUTARAM INOCORRENTE A APONTADA VIOLAÇÃO ANTE A INCIDÊNCIA AO CASO DA EXCEÇÃO LEGAL CONSTANTE DO ARTIGO 46, INCISO III, DA LEI Nº 9.610/98 - INSURGÊNCIA DO DEMANDANTE - RECLAMO DESPROVIDO.

Hipótese: Discussão afeta à aventada violação de direito autoral pela citação/reprodução, sem prévia autorização, da integralidade de um poema no bojo de determinado livro didático.

1. Incidência do óbice da Súmula 284/STF relativamente ao apontado malferimento do artigo 535 do CPC/73, ante a deficiência nas razões recursais dada a ausência de demonstração acerca da existência de defeito específico inerente ao acórdão embargado ensejador da alegada negativa de prestação jurisdicional.

2. As instâncias ordinárias compreenderam que o poema "O lagarto medroso" não seria em si considerado uma obra, visto que parte integrante do livro "Ou isto ou aquilo", no qual a poetisa Cecília Meireles aglutinou diversos textos poéticos, motivo pelo qual a citação da integralidade desse elemento textual não ensejaria violação ao direito autoral, por constituir mera passagem de fração da obra literária maior (o livro).

2.1 Apesar do poema ser parte integrante do livro, goza de proteção autoral nos termos da legislação de regência, pois consoante disposto no artigo 7º, caput e inciso I da Lei 9.610/98, "são obras intelectuais protegidas as criações de espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível", tais como os "textos de obras literárias". A poesia se constitui em uma criação do espírito expressa em forma de poema (gênero textual dividido em estrofes e versos), ou seja, é um texto de obra literária e a circunstância de ter sido publicado conjuntamente com outros não desnatura a sua qualificação, haja vista ser viável a coexistência da obra literária singular (poema) e da obra literária global (livro).

2.2 A despeito disso, a inadequação classificatória expressa pelas instâncias ordinárias não enseja o acolhimento da apontada violação ao direito autoral, ainda que tenha havido a reprodução, na íntegra, do texto "O lagarto medroso", porquanto além de ter sido indicado o nome do autor e a fonte bibliográfica, o legislador estabeleceu no artigo 46, inciso III da Lei 9.610/98, não constituir ofensa ao direito autoral a citação de "passagens de qualquer obra para fins de estudo, crítica ou polêmica, na medida justificada para o fim a atingir".

2.3 No caso, é incontroverso dos autos que o contexto de citação do poema no livro didático está revestido de padrões ligados à difusão educacional (ensino) e a "medida justificada para o fim a atingir", constitui e exige o estudo integral do poema, dele extraindo a análise ortográfica e semântica, procedimentos específicos para o ensino da língua portuguesa, mas também vislumbradas diversas outras análises interdisciplinares relacionadas à botânica, biologia, geografia, matemática (quantidade de estrofes, versos, repetições de palavras e rimas) entre outros.

2.4 A citação apenas parcial de trechos/excertos/passagens da poesia certamente prejudicaria a compreensão da criação intelectual da célebre poetisa Cecília Meireles e poderia gerar uma deturpação semântica do escrito, o que causaria a violação da integridade da obra literária e, conseqüentemente, ocasionaria atentado ao direito moral do autor consoante previsto no artigo 24, inciso IV da Lei 9.610/98.

3. Recurso especial conhecido em parte e, na extensão, desprovido." (Brasil, 2020b).

De qualquer forma, a ausência de uma casuística jurídica específica para a mineração de textos e dados nos leva a considerar o que já foi dito em relação ao modelo de cláusula aberta: ele constitui uma alternativa à via legislativa, mas não a obsta. Na verdade, como veremos a seguir, é justamente fora do Poder Judiciário que se parece situar o principal palco dos debates atuais sobre mineração de textos e dados no Direito Autoral brasileiro, incluindo seu papel na inovação e nas tecnologias de IA, bem como de que forma as questões trazidas pelas IA Generativa impactam na regulação dessas atividades.

5.2. Mineração de textos e dados na política brasileira de inovação

O valor das tecnologias intensivas em dados tem recebido, em anos recentes, uma atenção notadamente crescente por parte das políticas nacionais de inovação. Em 2019, o Ministério da Ciência, Tecnologia, Inovações e Comunicações, à época comandado pelo Ministro Marcos Pontes direcionava seus esforços para o desenvolvimento de tecnologias de inteligência artificial e Internet das Coisas no Brasil, com enfoque, entre outros pontos, na regulação no setor de ciência e tecnologia.²²² Isto culminou no *Plano Nacional de Internet das Coisas*, que visava o estímulo ao desenvolvimento dessa tecnologia no Brasil, pautando-se na livre concorrência e circulação de dados como alguns de seus principais alicerces (art. 1º).²²³ Frisa-se, porém, que o projeto não mencionava explicitamente a propriedade intelectual ou o diálogo com instituições de pesquisa e ensino naquele momento, debruçando-se sobre aspectos de cunho empresarial.

Isto mudaria alguns anos mais tarde, com a Estratégia Brasileira de Inteligência Artificial, instituída pela Portaria MCTI nº 4.617/2021 e depois alterada pela Portaria MCTI nº 4.979/2021 (Brasil, 2021c). A iniciativa reconhecia uma rápida expansão das tecnologias de aprendizado de máquina e a necessidade de atentar para suas implicações em diversas áreas, incluindo pesquisa, desenvolvimento e inovação.

Mais precisamente, dois pontos colocados pela Estratégia chamam a atenção: o primeiro é a seção de “PESQUISA, DESENVOLVIMENTO, INOVAÇÃO E EMPREENDEDORISMO”, que faz menção explícita ao compromisso de estimular a pesquisa e desenvolvimento de sistemas de IA, bem como a criação de parcerias entre setor público, privado e instituições de pesquisa como ações estratégicas a serem adotadas.²²⁴ O segundo ponto de destaque da

²²² Para mais detalhes, ver: <https://www.youtube.com/watch?v=DkWFfkCVHFw>. Acesso em 23 out. 2019.

²²³ “Art. 1º Fica instituído o Plano Nacional de Internet das Coisas com a finalidade de implementar e desenvolver a Internet das Coisas no País e , com base na livre concorrência e na livre circulação de dados, observadas as diretrizes de segurança da informação e de proteção de dados pessoais” (Brasil, 2019b).

²²⁴ “Acões Estratégicas

estratégia é, justamente, o reconhecimento não só da mineração de textos e dados como vital para o desenvolvimento de sistemas de IA, mas também das discussões, em âmbito internacional, sobre a necessidade de discutir uma limitação de direitos autorais específica para a mineração de textos e dados e de trazer este mesmo debate para o Brasil:

Estruturas regulatórias sobre a proteção e aplicação dos direitos de propriedade intelectual (DPI) também são importantes. Como os dados de treinamento para modelos de IA geralmente precisam ser copiados e editados, a clareza das regulamentações relacionadas à proteção de direitos autorais se mostra essencial. Um dos pontos de destaque neste tópico diz respeito à *necessidade de se incluir um novo tipo de limitação aos direitos autorais, para mineração de textos e de dados (Text & Data Mining exception)*. Esse foi um dos temas abordados na Segunda Sessão da Conferência entre Propriedade Intelectual e Inteligência Artificial da Organização Mundial de Propriedade Intelectual. Como os conjuntos de dados são organizados e processados, as regras existentes relacionadas à proteção de bancos de dados também podem desempenhar papel de destaque. (Brasil, 2021c, p. 18-19. Destaque nosso).

Além disso, Estratégia Nacional de Propriedade Intelectual, publicada em 2021, não só prevê, em seu artigo 1º, o uso da PI enquanto política pública de incentivo ao desenvolvimento e seu equilíbrio com a livre concorrência e interesse social, como, em diversos pontos (a exemplo de seu Eixo 6), prevê o uso da mineração de dados ou da construção de bancos de dados como ferramenta de auxílio do governo ao setor industrial e de tecnologia no Brasil (Brasil, 2021b).

5.3. Mineração de textos e dados e Direito Autoral no âmbito legislativo

No que tange à regulamentação da mineração de textos e dados, o Brasil se encontra em uma posição peculiar. Primeiramente, porque já possui o benefício de não ser o primeiro país a versar sobre a questão: como visto anteriormente, diferentes modelos regulatórios sobre o tema já existem em alguma medida no Direito Autoral em diversos países – seja tratando especificamente do uso em processos de análise de dados, seja inserindo-os em uma L&E mais ampla para pesquisa. Isto levaria à impressão de que o legislador se encontraria na cômoda posição de observar tais modelos e compará-los à realidade local.

Contudo, boa parte dos modelos (legislativos ou judiciais) vistos neste trabalho antecede a intensificação mais recente dos debates sobre IA Generativa ocorrida em 2022: mesmo Singapura adicionou sua regulação acerca da coleta e análise de dados antes disso, em 2021. Ao que tudo indica, este será um tópico a ser enfrentado pelo legislador, uma vez que a demanda

(...)

- Ampliar as possibilidades de pesquisa, desenvolvimento, inovação e aplicação de IA, por meio da viabilização do aporte de recursos específicos para esse tema e da coordenação entre iniciativas já existentes.

- Estabelecer conexões e parcerias entre setor público, setor privado e instituições científicas e universidades em prol do avanço no desenvolvimento e utilização da IA no Brasil.” (Brasil, 2021c, p. 37).

pela regulação dessa tecnologia perpassa o uso de obras no treinamento de sistemas de IA²²⁵ – o que, como também já observado, envolve o uso de técnicas de mineração de textos e dados.

Ainda assim, isto não significa exatamente que seja preciso pensar em uma proibição e, ao mesmo tempo, em uma limitação para a mineração de textos e dados. Afinal, de forma também peculiar, o Brasil já lista a inclusão de obras em bases de dados como um direito exclusivo do titular. Ora, se essas técnicas dependem da coleta de material e sua organização em um repositório, então pode-se entender que, na ausência de uma limitação específica para a pesquisa, não só a IA Generativa como qualquer outro projeto que dependa de obras protegidas para a construção de um *corpus* já se encontra regulado pelo Direito Autoral brasileiro desde 1998. Parece, portanto, que o contexto institucional brasileiro já opera além das considerações de Senftleben ao efetivamente já incluir (e restringir) a mineração de textos e dados dentro do domínio do Direito Autoral, evidenciando a necessidade de se pensar em uma regulação específica para a coleta e análise de obras protegidas na investigação científica – e como a inteligência artificial e seus impactos sociais podem influenciá-la.

Nesse contexto, os últimos anos testemunharam a proposição de diversos projetos de lei no Congresso Nacional que abordaram, em alguma medida, os sistemas de IA. Mais especificamente, o relatório da Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA) de 07 de junho de 2024 (Brasil, 2024g) aponta dez projetos que trataram do tema de forma específica, dos destaca-se o primeiro deles: o Projeto de Lei nº 21/2020, de autoria do Deputado Federal Eduardo Bismarck (PDT/CE), o qual visava a regulação do desenvolvimento e uso de sistemas de inteligência artificial no país de forma geral (Brasil, 2021c). Assim, o projeto não tratava especificamente de Direito Autoral, mas apresentava uma limitação voltada para a mineração de textos e dados sob a forma de um princípio geral para a criação e uso de IA no Brasil:

Art. 5o São princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil:

(...)

VIII – disponibilidade de dados: não violação do direito de autor pelo uso de dados, de banco de dados e de textos por ele protegidos, para fins de treinamento de sistemas de inteligência artificial, desde que não seja impactada a exploração normal da obra por seu titular. (Brasil, 2021a).

Tratava-se de um texto ao mesmo tempo restrito em certos pontos e bastante amplo em outros. A restritividade mais óbvia pode ser vista em relação à finalidade expressa do

²²⁵ Por exemplo, ver SCCR/44/8 (OMPI, 2023c).

dispositivo, que cobria apenas usos para fins de treinamento de IAs – algo à primeira vista apropriado, dado o teor do Projeto de Lei, mas que excluía, por exemplo, outros tipos de uso na pesquisa.

Entretanto, o texto ainda favorecia uma amplitude enorme de interpretação no seu escopo. Primeiramente, permitia-se o uso de obras protegidas de uma forma geral, sem especificar certos direitos exclusivos como reprodução ou comunicação ao público – na mesma tendência de alguns dos usos livres mais abertos na classificação de Flynn et al (2022). Em relação ao material, havia referência tanto às bases de dados em si como ao seu conteúdo, e a não-inteferência na “exploração normal da obra” tomava inspiração direta na regra dos três passos. Contudo, não havia proibição à interferência nos legítimos interesses do titular, tornando questionável o cumprimento do art. 9 (2) da Convenção de Berna.

Em 2022, foi formada a Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA) (Brasil, 2022a, 2022b), que conduziu um processo de consulta pública acerca da regulação da inteligência artificial no país. Realizaram-se audiências, um seminário e o recebimento de contribuições por escrito – os quais culminaram em um relatório final, publicado em dezembro de 2022 (Brasil, 2022c). Na seara da mineração de textos e dados, estudiosos se manifestaram nas audiências públicas e no seminário internacional no sentido de garantir segurança jurídica à prática, especificamente através da criação de limitações e exceções já implementadas em outros países (Ibid., p. 127; 197). Já as contribuições escritas foram mais diversas em conteúdo: múltiplos atores da sociedade civil expressaram a necessidade de um dispositivo legal para a mineração de textos e dados, enquanto representantes de titulares de direitos autorais (no caso, as Entidades dos Setores Editorial, Musical e Audiovisual e a Motion Pictures Association Brasil) criticaram o teor excessivamente amplo da limitação em trâmite naquele momento. Outros atores se manifestaram no sentido de que a pauta não deveria ser tratada naquele momento, enquanto a Fundação Getúlio Vargas colocou a necessidade de se esclarecer o que contaria como “exploração normal da obra” (Ibid., p. 257-259).

Feitas estas considerações e terminada a análise na Comissão, propôs-se um modelo de Substitutivo (Brasil, 2022c), que culminou na proposição do Projeto de Lei nº 2.338/2023, de autoria do Senador Rodrigo Pacheco (PSD/MG) (Brasil, 2024b). O texto legal contou com uma revisão profunda, que incluiu uma limitação para mineração de textos e dados com uma redação mais completa, detalhada e específica no seu escopo, que deixou de ser um princípio geral para ser dividido em duas partes. A primeira traz uma definição legal de mineração de

textos e dados no art. 4º, enquanto a segunda traz uma limitação para mineração de textos e dados na pesquisa, manifesta no art. 42 do Projeto:

Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

(...)

VIII – mineração de textos e dados: processo de extração e análise de grandes quantidades de dados ou de trechos parciais ou integrais de conteúdo textual, a partir dos quais são extraídos padrões e correlações que gerarão informações relevantes para o desenvolvimento ou utilização de sistemas de inteligência artificial.

(...)

Art. 42. Não constitui ofensa a direitos autorais a utilização automatizada de obras, como extração, reprodução, armazenamento e transformação, em processos de mineração de dados e textos em sistemas de inteligência artificial, nas atividades feitas por organizações e instituições de pesquisa, de jornalismo e por museus, arquivos e bibliotecas, desde que:

I – não tenha como objetivo a simples reprodução, exibição ou disseminação da obra original em si;

II – o uso ocorra na medida necessária para o objetivo a ser alcançado;

III – não prejudique de forma injustificada os interesses econômicos dos titulares; e

IV – não concorra com a exploração normal das obras.

§ 1º Eventuais reproduções de obras para a atividade de mineração de dados serão mantidas em estritas condições de segurança, e apenas pelo tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados da pesquisa científica.

§ 2º Aplica-se o disposto no caput à atividade de mineração de dados e textos para outras atividades analíticas em sistemas de inteligência artificial, cumpridas as condições dos incisos do caput e do § 1º, desde que as atividades não comuniquem a obra ao público e que o acesso às obras tenha se dado de forma legítima.

§ 3º A atividade de mineração de textos e dados que envolva dados pessoais estará sujeita às disposições da Lei no 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais). (Brasil, 2022c).

Aqui, observa-se um texto que em muito lembra uma amálgama de modelos encontrados em outras limitações do gênero no mundo, sobretudo no Japão e na União Europeia.

As referências ao modelo japonês se dão, mais diretamente, no estabelecimento das condições para que a mineração de textos e dados no contexto da IA se observe, presentes nos incisos I a IV: há limitação quantitativa leve, “na medida necessária”, e uma proibição ao uso que leve à “reprodução, *exibição* ou disseminação” (destaque nosso), algo próximo ao disposto na lei japonesa quando é dada permissão a qualquer uso que não leve ao apreçamento sensorial da obra original, além da menção explícita à não-exploração normal da obra ou aos interesses legítimos do titular – esta última ausente na primeira versão do texto brasileiro (Japão, 2018).

Já a inspiração europeia é mais visível nos parágrafos do art. 42: não apenas há a cobrança de manutenção do corpus sob um regime de segurança – de forma a não permitir a apropriação e uso por terceiros com outras intenções – como também se observa a restrição de

que o acesso ao material precisa ter se dado de forma legítima. Note-se que este último ponto não constitui exatamente uma novidade no Direito Autoral brasileiro: o art. 30, §1º, ao permitir cópias transitórias, qualifica que esta deverá decorrer exclusivamente de uso autorizado pelo titular.²²⁶

Por outro lado, a limitação não se restringia a meramente reproduzir tais dispositivos, introduzindo alguns elementos peculiares: o primeiro é a sua extensão não somente para a investigação científica, mas também a jornalística. Trata-se de uma inclusão que não encontra paralelo em nenhum outro dispositivo do gênero abordado nesta pesquisa, e que apresenta pertinência na medida em que isto ajuda a legitimar o funcionamento de ferramentas de checagem de fatos – as quais podem utilizar sistemas automatizados, como mencionado no Capítulo 2. Também são incluídos museus, arquivos e bibliotecas, o que facilita a digitalização de acervos. É importante mencionar que a extensão das L&E para estes fins não constitui uma pauta recente no Direito Autoral: conforme mencionado no Capítulo 2, isto já é, inclusive, objeto de discussão na OMPI há anos.

Outro elemento peculiar diz respeito à delimitação do escopo da limitação: por um lado, a restrição para atividades voltadas estritamente ao treinamento de inteligência artificial permaneceu, inclusive em função de a definição de mineração de textos e dados adotada pelo substitutivo estar circunscrita ao uso por IA – posição esta que será discutida mais a fundo em um momento posterior. Por outro lado, destaca-se que o objeto da limitação no art. 42 também se estendia para além do simples treinamento, cobrindo também o uso da mineração de textos e dados por IA de forma geral. Isto se reforça no § 2º, que estende a hipótese de uso lícito também para qualquer análise de dados como um todo que faça uso de inteligência artificial – o que, dada a definição ampla de inteligência artificial no projeto e ao caráter cada vez mais ubíquo da coleta e análise automatizada de dados (como visto no Capítulo 1), poderia representar maior segurança para diversos setores que fazem uso não-expressivo de obras.

Em agosto de 2023, com o fim da Comissão de Juristas, foi requerida e criada a Comissão Temporária Interna sobre Inteligência Artificial no Brasil, que passou a ser a responsável por realizar reuniões e consultas sobre a regulação da IA no país (Brasil, 2023). Entre abril

²²⁶ “Art. 30.

(...)

§ 1º O direito de exclusividade de reprodução não será aplicável quando ela for temporária e apenas tiver o propósito de tornar a obra, fonograma ou interpretação perceptível em meio eletrônico ou quando for de natureza transitória e incidental, desde que ocorra no curso do uso devidamente autorizado da obra, pelo titular “(Brasil, 1998b).

e maio de 2024, a Comissão recebeu novas manifestações dos setores criativos: a primeira, de 08 de abril, assinada por diversas entidades de representação da classe artística e jornalística se manifestou contra a mineração de textos e dados e o desenvolvimento de inteligência artificial sem autorização, mantendo-se resistentes a qualquer limitação de direitos autorais, mas reforçando a regra dos três passos no caso de um dispositivo do gênero.²²⁷ A outra manifestação, de 07 de maio, assinada pelo SINDCINE, se restringiu a um pedido de audiência para maiores esclarecimentos acerca de suas preocupações sobre o tema.²²⁸

Em 24 de abril de 2024, a CTIA apresentou um Substitutivo Preliminar para o Projeto de Lei, o qual fez alterações profundas na regulação sobre Direito Autoral: tinha-se, agora, uma seção à parte dedicada aos direitos de autor e conexos, com diversas modificações. Dentre elas, incluiu-se a inclusão de uma obrigação de transparência, por parte de usuários e desenvolvedores, sobre quais conteúdos foram utilizados no treinamento de sistemas de IA; a inclusão do requisito de acesso legítimo à limitação para mineração de textos e dados e a remoção da permissão para atividades analíticas de IA, em favor de uma cláusula que exclui entidades controladas por empresas e, especialmente, a inclusão de cláusulas permitindo aos titulares proibir o uso de obras em treinamento de IA e estabelecer a cobrança de remuneração por este uso. O Substitutivo também remove as ressalvas feitas à LGPD e prevê um ambiente experimental (*sandbox*) regulatório, pelo qual a autoridade estatal competente estabelecerá critérios para a regulação de regras referentes à transparência e à remuneração pelo uso de obras em IA (Brasil, 2024f). Esta proposta mais tarde sofreria emendas e culminaria em novos modelos,

²²⁷ “Portanto, para garantir os direitos de propriedade intelectual, o PL 2338/2023 deve ter clara previsão para estabelecer que:

a) O uso de obras e produções protegidas para mineração de dados e desenvolvimento de ferramentas de IA deverá estar submetido à autorização prévia.

b) Os conteúdos gerados por IA não poderão ser assemelhados ou protegidos pelas normas da propriedade intelectual.

c) Apesar dos setores da indústria criativa não encontrarem espaço para novas exceções e limitações aos direitos autorais, caso a legislação venha a inserir alguma previsão neste sentido, as exceções e limitações da mineração de textos e dados e o desenvolvimento de ferramentas de IA deverão ser restritas e submetidas ao teste dos três passos previsto na lei brasileira e nos tratados internacionais, que regulam os princípios dos direitos autorais, sempre preservada a prerrogativa de o titular de direitos autorais autorizar ou proibir o uso de sua obra ou produção. As exceções nunca deverão ser consideradas ou utilizadas para a mineração de dados ou o treinamento comercial, e os usos não comerciais só deverão ser permitidos em determinadas situações estabelecidas por diretrizes claras e jamais admitir prejuízo injustificado aos titulares de direitos autorais.”

²²⁸ “É sabido, e já manifestado por V.Exa. em várias oportunidades, que a evolução extraordinária da IA já está impactando e impactará ainda mais todos os setores de nossa economia e de nossa vida social, daí a importância de uma regulamentação democrática que cerceie os abusos, proteja as instituições públicas e privadas e também não cerceie ou coloque barreiras ao desenvolvimento tecnológico.

Por conta desses impactos, já perceptíveis em nossa sociedade, e, reforçando, por não termos notado a existência de discussões sobre nosso setor, estamos solicitando a V. Exa. uma audiência para apresentarmos, para sua análise e avaliação, sob o ponto de vista dos trabalhadores de nosso setor, a exposição do problema, a nossa visão e pautar a discussão da aplicação da ferramenta no ambiente audiovisual, que já está sendo aplicada sem regulação.” (SINDCINE, 2024).

apresentados em dois pareceres apresentados ao Senado Federal em 07 e 18 de junho de 2024 (Brasil, 2024d; 2024e).

Até o dia 10 de dezembro de 2024, 244 emendas foram propostas ao projeto (Brasil, 2024c), algumas das quais contém propostas de alteração especificamente direcionadas para o uso de obras protegidas em mineração de textos e dados e/ou no treinamento de IA – cuja relação completa pode ser vista no Apêndice B deste trabalho. Mais especificamente, a Emenda nº 44 (aprovada), do Senador Chico Rodrigues (PSB/RR), visava alterar o agora art. 59 para impor a cláusula de transparência a desenvolvedores ao invés de aplicadores de IA, com a ideia de retirar a responsabilidade legal sobre os usuários – que, afinal, não são diretamente responsáveis pela construção e treino dos sistemas.²²⁹ Já a Emenda nº 45 (aprovada em parte) visou modificar a disposição sobre *sandbox* regulatório do art. 63, direcionando para o SAI a competência sobre o mesmo e, principalmente, removendo a ressalva a usos não-comerciais.

Mais recentemente, a Emenda nº 191 (aprovada), de Sen. Efraim Filho (UNIÃO/PB), estabelece complementos à nova cláusula de transparência presente no substitutivo, estipulando meios de efetivação do cumprimento deste dispositivo.²³⁰ Já a Emenda nº 192 (aprovada em parte), do mesmo autor, modifica a limitação presente no PL para deixar mais clara a permissão para atividades de mineração de textos e dados, ressaltar sua finalidade não-comercial e a proibição da disseminação de obras protegidas. Ao mesmo tempo, porém, remove o uso jornalístico do texto.²³¹ A Emenda nº 193 (aprovada) visa esclarecer melhor os meios para efetivação das prerrogativas de proibição do uso de obras em IA.

²²⁹ “A emenda proposta concentra toda a responsabilidade de informar sobre o uso de conteúdos protegidos exclusivamente no desenvolvedor. Com a responsabilidade limitada aos desenvolvedores, fica claro quem deve cumprir com as obrigações legais, eliminando ambiguidades que podem surgir quando múltiplos agentes estão envolvidos no ciclo de vida dos sistemas de IA. O desenvolvedor, sendo o "criador" do sistema, está em melhor posição para monitorar e relatar o uso de conteúdo protegido por direitos autorais e conexos. (...) Nesse contexto, indivíduos, organizações ou entidades que apenas utilizam os sistemas de IA desenvolvidos por terceiros, ou seja, os aplicadores, não têm a obrigação de informar sobre o conteúdo utilizado no treinamento desses sistemas. Isso simplifica a cadeia de responsabilidade, reduzindo os custos associados ao cumprimento de obrigações legais e incentivando a inovação de novas aplicações para IA em diversos setores, como saúde, educação, finanças e transporte.”

²³⁰ “JUSTIFICAÇÃO (...)

O art. 62 estabelece a obrigatoriedade de que os desenvolvedores publiquem, em sítio eletrônico de fácil acesso, um sumário de todos os conteúdos protegidos por direitos autorais e conexos utilizados durante o desenvolvimento de sistemas de inteligência artificial.

Essa medida garante transparência no uso dessas obras, permitindo que os titulares identifiquem como seus conteúdos estão sendo utilizados, permitindo-lhes negociar conforme a legislação vigente.

Ao especificar que o termo "desenvolvimento" engloba as etapas de mineração, treinamento, retreinamento, teste, validação e aplicação, o dispositivo assegura uma abordagem abrangente que cobre todo o ciclo de vida dos sistemas de inteligência artificial.” (Brasil, 2024c).

²³¹ “JUSTIFICAÇÃO

(...)

Para tal, estabelece: (i) que apenas conteúdos obtidos de forma legal possam ser utilizados, evitando apropriação indevida ou pirataria; (ii) limite do uso por atividades que não visem lucro, protegendo o mercado dos titulares e

Outras propostas de destaque são as emendas nº 24, do Senador Laércio Oliveira (PP/SE) (posteriormente retirada), e 61, do Senador Astronauta Marcos Pontes (PL/SP) (rejeitada), que visavam remover, no todo ou em parte, a seção sobre Direito Autoral do projeto. Ambas se posicionam, em suas justificativas, contrárias ao escopo da proteção criada em lei, alegando, de diversas maneiras, que ela imporá barreiras à inovação no Brasil: seja pela inviabilidade de efetivação,²³² seja por aumentar uma proteção legal que se entenderia já existente.²³³

Já outras propostas de emenda trazem redações alternativas para limitações e exceções: a Emenda nº 72, do Sen. Marcos Pontes, propõe um novo texto para a seção de Direito Autoral, com uma limitação inteiramente nova: abrangendo todos os direitos exclusivos (com a ressalva de só ser possível compartilhar as obras para verificação de resultados ou colaboração em pesquisa), sem restrições a fins comerciais.²³⁴ Já a Emenda nº 159, do Senador Carlos Portinho

restringindo a exceção às finalidades educacionais e científicas; e (iii) que o conteúdo protegido só pode ser utilizado na medida estritamente necessária, garantindo que os interesses econômicos dos titulares não sejam prejudicados e que não haja competição direta com a exploração comercial das obras.

Ressalta-se a exclusão do “jornalismo” da lista de exceções previstas no art. 63, considerando sua natureza predominantemente comercial ou economicamente explorável, o que o diferencia das demais atividades contempladas, como as de museus, bibliotecas e instituições educacionais, que têm foco essencialmente na preservação histórica, educação e pesquisa científica.

Além disso, instituições acadêmicas da área de comunicação podem realizar pesquisas envolvendo mineração de textos e dados, uma vez que continuam abrangidas pelo dispositivo.

Os parágrafos apresentados visam regulamentar com maior precisão o uso de conteúdos protegidos por direitos autorais em sistemas de IA estabelecendo salvaguardas para os titulares de direitos, enquanto permitem o uso legítimo em contextos específicos.” (Brasil, 2024c).

²³² Conforme justifica a Emenda nº 44: “(...) A exigência de informar quais conteúdos protegidos por direitos autorais foram utilizados (Art. 59) e a restrição de uso comercial (Inciso II do Art. 60) impõem barreiras significativas para desenvolvedores, dificultando o acesso a dados essenciais (...). No processo de machine learning, são utilizados centenas, milhares e, às vezes, milhões de conteúdos distintos. A exigência de identificação e obtenção de consentimento do titular para cada um desses conteúdos torna o processo inviável, criando uma barreira intransponível para a inovação (...). Com as restrições impostas pelos dispositivos atuais, os modelos de IA podem não refletir de maneira precisa as nuances e particularidades do Brasil, atrasando o uso dessas tecnologias no país. Isso inclui a capacidade de combater conteúdo discriminatório de forma eficaz (...).” (Brasil, 2024c).

²³³ Segundo o texto da Emenda nº 45: “A inclusão de tais dispositivos neste projeto de lei, que trata da regulamentação da IA como um todo, pode gerar inseguranças jurídicas e prejudicar o desenvolvimento tecnológico do país. (...)

Além disso, a legislação autoral brasileira já possui mecanismos para proteger os direitos dos autores em relação ao uso de suas obras por sistemas de IA. A Lei de Direitos Autorais (Lei nº 9.610/1998). Esses mecanismos já oferecem proteção adequada aos titulares de direitos autorais, tornando desnecessária a inclusão de novas disposições no PL 2338/2023.” (Brasil, 2024c).

²³⁴ “Art. 60. O desenvolvedor de um sistema de inteligência artificial poderá utilizar obras protegidas por direito autoral e conteúdos relacionados no desenvolvimento, desde que sejam cumpridas as seguintes condições:

I – o uso da obra esteja limitado a:

- a) análise computacional de dados; ou
- b) preparação da obra ou gravação para análise computacional de dados;

II – o desenvolvedor não utilize a obra para qualquer outro fim;

III – o desenvolvedor não compartilhe uma cópia da obra com terceiros, exceto para fins de:

- a) verificação dos resultados da análise computacional de dados realizada pelo fornecedor; ou
- b) pesquisa colaborativa ou estudo relacionado ao propósito da análise computacional de dados realizada pelo fornecedor;

IV – o desenvolvedor tenha acesso legal ao material utilizado no desenvolvimento e/ou o material esteja disponibilizado licitamente de forma gratuita na internet para acesso por qualquer pessoa.”

(PL/RJ), traz uma limitação para mineração de textos e dados em múltiplos artigos, com destaque notável para a pesquisa e com um texto que traz certa semelhança à limitação presente na legislação alemã – vista no Capítulo 3.²³⁵ O texto desta última Emenda também é adotado como base para as Emendas nº 161, do Sen. Mecias de Jesus (REPUBLICANOS/RR), e 187, do Sen. Izalci Lucas (PL/DF) (parcialmente aprovada). Já as Emendas nº 162 e 168, também de autoria do Sen. Carlos Portinho, mantêm o texto da limitação do substitutivo, porém impondo-lhe uma reserva legal nos moldes do art. 04 da Diretiva CDSM.²³⁶ Tal proposta é replicada na Emenda nº 185, do Sen. Izalci Lucas (parcialmente aprovada).

Avaliadas as propostas de emenda até então, foi publicada uma Complementação de Voto em 05 de dezembro de 2024, com a versão final do Projeto de Lei dentro do Senado.²³⁷ O modelo atual da regulação sobre IA e Direito Autoral no PL 2.338/2023 agora ocupa a Seção

²³⁵ Seção IV

Direitos de autor e conexos

Art. 62. O desenvolvedor de inteligência artificial poderá utilizar obras protegidas por direitos de autor e conexos no seu desenvolvimento ou configuração, sujeito às seguintes condições:

I - A utilização da obra não tenha como objetivo principal a reprodução, exibição ou disseminação da própria obra original;

II - O desenvolvedor ou aplicador não compartilhe uma cópia da obra com quaisquer terceiros exceto para fins de (a) verificação dos resultados da análise de dados computacionais realizada pelo desenvolvedor; ou (b) pesquisa colaborativa ou estudos relacionados às finalidades da análise de dados computacionais realizada pelo desenvolvedor;

III - O desenvolvedor ou aplicador tenha legalmente acessado o material usado no desenvolvimento.

Art. 63. Não constitui ofensa aos direitos de autor e conexos a utilização automatizada de conteúdos protegidos em processos de mineração de textos e dados para os fins de pesquisa ou desenvolvimento de sistemas de inteligência artificial sujeito às seguintes condições:

I - Eventuais cópias de conteúdos protegidos por direitos de autor e conexos utilizadas no treinamento de sistemas de inteligência artificial deverão ser mantidas em estritas condições de segurança, e unicamente pelo tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados.

II - A atividade de mineração de textos e dados que envolva dados pessoais estará sujeita às disposições da Lei no 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais).

Art. 64. O titular de direitos de autor e conexos poderá optar por reservar seus direitos sobre as obras para impedir a mineração de textos e dados para o desenvolvimento de sistemas de inteligência artificial, exceto quando esta for feita para fins de pesquisa científica, e desde que esta reserva seja realizada através de métodos compreensíveis por sistemas computacionais.

Parágrafo único. Nos casos em que a reserva de direitos tiver sido realizada expressa e validamente de forma adequada, e quando nenhuma outra exceção de uso seja aplicável, o desenvolvedor e o aplicador de um sistema de inteligência artificial precisarão obter autorização dos titulares de direitos autorais e conexos para a realização de mineração de texto e dados sobre as obras.

²³⁶“Art. 62. Não constitui ofensa aos direitos autorais a utilização automatizada de conteúdos protegidos em processos de mineração de textos e dados para os fins de pesquisa ou desenvolvimento de sistemas de inteligência artificial, desde que observadas as seguintes condições:

(...)

IV - não exista sinalização expressa junto a obra da oposição a sua análise para o desenvolvimento de sistemas de inteligência artificial, devendo tal oposição ser feita por meio legível por máquina e padronizado quando a obra for disponibilizada na rede mundial de computadores, nos termos desta lei e da regulação.” (Brasil, 2024c).

²³⁷ Para uma comparação mais detalhada entre as diferentes versões dos dispositivos sobre direitos autorais no PL, ver Instituto Brasileiro de Direitos Autorais (2024c).

IV (“Dos Direitos de Autor e Conexos”) do Capítulo IX do referido texto (“DO FOMENTO À INOVAÇÃO SUSTENTÁVEL”), descrita a seguir:

Seção IV

Dos Direitos de Autor e Conexos

Art. 62. O desenvolvedor de IA que utilizar conteúdo protegido por direitos de autor e conexos deverá informar sobre os conteúdos protegidos utilizados nos processos de desenvolvimento dos sistemas de IA, por meio da publicação de sumário em sítio eletrônico de fácil acesso, observados os segredos comercial e industrial, nos termos de regulamento específico.

Parágrafo único. Para fins deste Capítulo, o desenvolvimento compreende as etapas de mineração, treinamento, retreinamento, testagem, validação e aplicação de sistemas de IA.

Art. 63. Não constitui ofensa aos direitos de autor e conexos a utilização automatizada de conteúdos protegidos em processos de mineração de textos e dados para os fins de pesquisa e desenvolvimento de sistemas de IA por organizações e instituições científicas, de pesquisa e educacionais, museus, arquivos públicos e bibliotecas, desde que observadas as seguintes condições:

I – o acesso tenha se dado de forma lícita;

II – não tenha fins comerciais;

III – a utilização de conteúdos protegidos por direitos de autor e conexos seja feita na medida necessária para o objetivo a ser alcançado, sem prejuízo dos interesses econômicos dos titulares e sem concorrência com a exploração normal das obras e conteúdos protegidos.

§ 1º Cópias de conteúdos protegidos por direitos de autor e conexos utilizadas nos sistemas de IA deverão ser armazenadas em condições de segurança, e unicamente pelo tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados.

§ 2º É vedada a exibição ou a disseminação das obras e conteúdos protegidos por direitos de autor e conexos utilizados no desenvolvimento de sistemas de IA.

§ 3º Este artigo não se aplica a instituições vinculadas, coligadas ou controladas por entidade com fins lucrativos que forneçam sistemas de IA ou que tenham, entre elas, participação acionária.

§ 4º Aplica-se o disposto no caput deste artigo à mineração de dados, por entidades públicas ou privadas, no contexto de sistemas de IA para combate a ilícitos civis e criminais, que atentem contra direitos de autor e conexos.

Art. 64. O titular de direitos de autor e conexos poderá proibir a utilização dos conteúdos de sua titularidade no desenvolvimento de sistemas de IA nas hipóteses não contempladas pelo art. 63 desta Lei.

Parágrafo único. A proibição do uso de obras e conteúdos protegidos nas bases de dados de um sistema de IA posterior ao processo de treinamento não exime o agente de IA de responder por perdas e danos morais e materiais, nos termos da legislação aplicável.

Art. 65. O agente de IA que utilizar conteúdos protegidos por direitos de autor e conexos em processos de mineração, treinamento ou desenvolvimento de sistemas de IA deve remunerar os titulares desses conteúdos em virtude dessa utilização, devendo-se assegurar:

I – que os titulares de direitos de autor e conexos tenham condições efetivas de negociar coletivamente, nos termos do Título VI da Lei nº 9.610, de 19 de fevereiro de 1998 (Lei dos Direitos Autorais), ou diretamente a utilização dos conteúdos dos quais são titulares, podendo fazê-lo de forma gratuita ou onerosa;

II – que o cálculo da remuneração a que se refere o caput considere os princípios da razoabilidade e da proporcionalidade e elementos relevantes, tais como o porte do agente de IA e os efeitos concorrenciais dos resultados em relação aos conteúdos originais utilizados;

III – a livre negociação na utilização dos conteúdos protegidos, visando à promoção de ambiente de pesquisa e experimentação que possibilite o desenvolvimento de práticas inovadoras, e que não restrinjam a liberdade de pactuação entre as partes envolvidas, nos termos dos arts. 156, 157, 421, 422, 478 e 479 da Lei nº 10.406, de 10 de janeiro de 2002 (Código Civil), e o art. 4º da Lei nº 9.610, de 19 de fevereiro de 1998 (Lei dos Direitos Autorais).

§ 1º A remuneração a que se refere o caput deste artigo é devida somente:

I – aos titulares de direitos de autor e conexos nacionais ou estrangeiros domiciliados no Brasil;

II – a pessoas domiciliadas em país que assegure a reciprocidade na proteção, em termos equivalentes a este artigo, aos direitos de autor e conexos de brasileiros, conforme disposto nos arts. 2º, parágrafo único, e 97, § 4º, da Lei nº 9.610, de 19 de fevereiro de 1998 (Lei dos Direitos Autorais), sendo vedada a cobrança nos casos em que a reciprocidade não estiver assegurada.

§ 2º O titular do direito de remuneração previsto no caput que optar pela negociação e autorização direta, nos termos do inciso I do caput, poderá exercê-lo independentemente de regulamentação posterior.

Art. 66. A utilização de conteúdos de imagem, áudio, voz ou vídeo que retratem ou identifiquem pessoas naturais pelos sistemas de IA deverá respeitar os direitos da personalidade, na forma prevista na Lei nº 10.406, de 10 de janeiro de 2002 (Código Civil), e na legislação pertinente. (Brasil, 2024b, p. 30-32).

O texto do substitutivo não difere das versões anteriores apenas em tamanho e forma: seu conteúdo parece refletir, em grande medida, questões trazidas pela IA Generativa, bem como regulações mais recentes, como o *AI Act* da União Europeia. Também traz disposições novas e muito mais restritivas que as versões anteriores, denotando um espírito de expansão que em muito lembra o que fora visto nos anos 90 – conforme visto em seções anteriores.

A inspiração mais imediata do caso europeu, em especial o *AI Act*, parece ser o requerimento de transparência previsto no art. 62. Por este artigo, qualquer obra utilizada no treinamento de um sistema de IA precisa ser identificada e sua inclusão, abertamente acessível para conferência. À primeira vista, isto significa que pesquisadores desenvolvendo ferramentas nesse sentido terão a responsabilidade de documentar detalhadamente cada obra utilizada e identificar quais são protegidas e quais não são. Os critérios específicos para tanto foram deixados a cargo de regulação posterior.

Já a limitação de Direito Autoral do art. 63 chama atenção porque traz alterações importantes em relação à sua versão anterior. A primeira é a extensão da limitação para atividades com pesquisa para além do simples treinamento de sistemas de inteligência artificial, acompanhada de uma alteração no parágrafo segundo que remove a permissão para atividades analíticas com inteligência artificial para além do seu treinamento: agora, o parágrafo explicitamente

remove da limitação entidades com fins lucrativos, ou que tenham alguma ligação institucional com estas. Trata-se de um dispositivo que cobre situações como as da OpenAI, que lida com IA Generativa e se declara como não-comercial, embora assuma possuir uma subsidiária com fins lucrativos como forma de complementar suas fontes de manutenção financeira (OpenAI, 2024e).²³⁸

O art. 64, por sua vez, constitui a primeira parte das prerrogativas agora conferidas pelo PL aos titulares de direitos autorais. Mais especificamente, este artigo em particular garante ao titular a possibilidade de proibir o uso de obras protegidas para fins de treinamento de sistemas de IA que não se enquadre nos usos especificados no artigo anterior – com uma proibição adicional, no art. 63, a tratamentos discriminatórios por parte de plataformas em relação a titulares que imponham estes termos. Trata-se de uma proibição redundante à primeira vista: conforme exposto anteriormente, o art. 29, IX da Lei 9.610/98 já torna obrigatório o consentimento do titular para a inclusão de uma obra em base de dados – o que representa um procedimento padrão para a mineração de textos e dados e, portanto, para o desenvolvimento de inteligências artificiais. Entretanto, é importante lembrar a possibilidade de que sistemas de IA façam uso de bancos de dados de terceiros como base para sua pesquisa, o que poderia representar uma fuga da proibição presente na LDA.

O art. 65, por sua vez, estipula a outra metade dessas prerrogativas ao estabelecer que ao titular é devida remuneração pelo uso de obras protegidas na mineração, treinamento ou desenvolvimento de sistemas de IA, com uma série de critérios a serem adotados pelas partes envolvidas na negociação e implementação dessa remuneração. Note-se que o estabelecimento destes parâmetros, antes inserido na esfera de competência do *sandbox* regulatório em versões anteriores do projeto, foi removido na versão atual, ficando inteiramente a cargo das partes envolvidas no caso concreto. Também cabe notar que a obrigação do atual art. 65, diferente da cláusula de transparência do art. 62, estende-se não apenas aos desenvolvedores, mas a todos os agentes de inteligência artificial – o que, pelo art. 4º, VII e VIII do PL, inclui distribuidores e aplicativos, ou usuários, dos sistemas de IA.²³⁹

²³⁸ “It became increasingly clear that donations alone would not scale with the cost of computational power and talent required to push core research forward, jeopardizing our mission. So we devised a structure to preserve our Nonprofit’s core mission, governance, and oversight while enabling us to raise the capital for our mission:

- The OpenAI Nonprofit would remain intact, with its board continuing as the overall governing body for all OpenAI activities.
- A new for-profit subsidiary would be formed, capable of issuing equity to raise capital and hire world class talent, but still at the direction of the Nonprofit. Employees working on for-profit initiatives were transitioned over to the new subsidiary.” (OpenAI, 2024e).

²³⁹ “Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

Aqui, diferentemente de países como EUA e Japão (que se concentram sobre o *output* do uso), o PL propõe um sistema de proibição e remuneração pelo uso de obras na fase de *input*, no treinamento em si – o que, para Frosio, seria mais condizente com a teoria geral do Direito Autoral no que diz respeito à exclusividade da proteção, especialmente tendo-se em mente o argumento de que a IA Generativa só seria possível pelo uso de obras prévias e o fato de que os desenvolvedores não seriam responsáveis pelo *output*, o qual já seria coberto pelas hipóteses padrões de violação na lei (Frosio, 2023, p. 19). Também se trata de um dispositivo legal teoricamente compatível com o rol de direitos exclusivos da LDA, que garante ao autor o controle sobre quaisquer modalidades de uso existentes ou futuras (art. 29, X). Por outro lado, a fim de evitar problemas de sobreinclusão da norma, entendemos que esta escolha necessita de maiores considerações acerca do seu escopo – as quais discutiremos mais adiante.

Já o artigo 66 se refere ao respeito aos direitos de personalidade contidos na aparência ou da voz de pessoas físicas por sistemas de IA. Trata-se de uma pauta cuja inclusão na seção de direitos de autor do PL 2.338/2023 parece, à primeira vista, desconexa do Direito Autoral, e mais próxima do direito de imagem no Código Civil. No entanto, o dispositivo ganha relevância em face de demandas sobre o uso não-autorizado da imagem de artistas nas greves de atores dos EUA – conforme abordado no Capítulo 1. Ademais, é coerente com os direitos conexos garantidos a intérpretes e executantes em relação à sua voz e imagem, mesmo de forma póstuma,²⁴⁰ que se opõem à garantia de continuidade da exploração econômica póstuma no art. 92

VII - aplicador: pessoa natural ou jurídica, de natureza pública ou privada, que empregue ou utilize, em seu nome ou benefício, sistema de inteligência artificial, inclusive configurando, gerenciando, mantendo ou apoiando com o fornecimento de dados para a operação e o monitoramento do sistema de IA;

VIII - agentes de inteligência artificial: desenvolvedores, distribuidores e aplicadores que atuem na cadeia de valor e na governança interna de sistemas de inteligência artificial, nos termos definidos por regulamento;" (Brasil, 2024b).

²⁴⁰ "Art. 90. Tem o artista intérprete ou executante o direito exclusivo de, a título oneroso ou gratuito, autorizar ou proibir:

I - a fixação de suas interpretações ou execuções;

II - a reprodução, a execução pública e a locação das suas interpretações ou execuções fixadas;

III - a radiodifusão das suas interpretações ou execuções, fixadas ou não;

IV - a colocação à disposição do público de suas interpretações ou execuções, de maneira que qualquer pessoa a elas possa ter acesso, no tempo e no lugar que individualmente escolherem;

V - qualquer outra modalidade de utilização de suas interpretações ou execuções.

(...)

§ 2º A proteção aos artistas intérpretes ou executantes estende-se à reprodução da voz e imagem, quando associadas às suas atuações.

(...)

Art. 92. Aos intérpretes cabem os direitos morais de integridade e paternidade de suas interpretações, inclusive depois da cessão dos direitos patrimoniais, sem prejuízo da redução, compactação, edição ou dublagem da obra de que tenham participado, sob a responsabilidade do produtor, que não poderá desfigurar a interpretação do artista. Parágrafo único. O falecimento de qualquer participante de obra audiovisual, concluída ou não, não obsta sua exibição e aproveitamento econômico, nem exige autorização adicional, sendo a remuneração prevista para o falecido, nos termos do contrato e da lei, efetuada a favor do espólio ou dos sucessores." (Brasil, 1998b).

da LDA, caso sejam utilizadas imagens do falecido sob controle do titular para fins de treinamento de IA.

Com estas alterações mais recentes ao texto do Projeto, observou-se uma mudança notável na postura observada em alguns grupos de interesse: diversas das entidades de representação artística que antes manifestavam reticência a uma limitação para MTD, agora mais explicitamente defendiam os parâmetros da limitação agora existente.²⁴¹ Há lógica neste movimento, uma vez que o modelo atualmente proposto possui restrições adicionais que denotam a busca por uma resposta estatal aos anseios dos titulares de direitos de autor e conexos.

Assim, o debate regulatório corrente sobre mineração de textos e dados, assim como em outras partes do mundo, parece ter se direcionado para a questão da inteligência artificial – movimento este que, diante do alcance destas tecnologias na vida social, na economia e na pesquisa, demanda uma análise mais detida no tocante à eficácia e exequibilidade da regulação atualmente em discussão, bem como seus efeitos sobre as atividades de pesquisadores, jornalistas e entidades de conservação da memória.

5.4. Principais questionamentos

5.4.1. Escopo dos sistemas de IA a serem regulamentados

A primeira questão que se coloca é o escopo do que exatamente conta como inteligência artificial para os efeitos da regulamentação de Direito Autoral pretendida pelo Legislativo. Nesse sentido, observa-se a adoção de uma conceituação ampla, que visa regulamentar diversos tipos de sistema – cujo exemplo mais recente pode ser visto no PL nº 2.338/2023:

Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

I - sistema de inteligência artificial (IA): sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial, previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real; (Brasil, 2024b).

²⁴¹ “É imperioso garantir que os titulares de direitos autorais tenham conhecimento do uso de suas obras, por meio da identificação e informação aos seus detentores pelos aplicativos e desenvolvedores de sistema de inteligência artificial generativa, e que as limitações e exceções propostas fiquem restritas às entidades de pesquisa, jornalismo, museus, arquivos, bibliotecas e educacionais, mediante às observações constantes desde que cumpridos os requisitos constantes do artigo 61 do substitutivo.

(...)

Assim, as organizações que subscrevem essa nota pedem aos senadores e senadoras que aprovem o PL nº 2338/2023 na Comissão Temporária de Inteligência Artificial, com as previsões de direitos autorais constantes do substitutivo visando o respeito aos direitos de criadores de conteúdos artísticos, intelectuais e jornalísticos.” (Associação de Jornalismo Digital, 2024).

Trata-se de uma abordagem, a princípio, lógica – uma conceituação ampla ajuda a evitar o defasamento do texto legal diante de desenvolvimentos tecnológicos – e que já atendia a comentários vistos no relatório final da CJSUBIA, que predominantemente criticavam uma excessiva restritividade do termo em projetos anteriores (Brasil, 2022c, p. 580-595). No entanto, é importante frisar que a aplicação de uma interpretação ampla em um contexto específico, como o de Direito Autoral, pode incluir tecnologias que, embora façam uso de algoritmos para coletar e interpretar dados que envolvem obras protegidas a fim de gerar conteúdo ou recomendações, não constituem IA Generativa, tampouco concorrem com as obras ou titulares de direitos. Um exemplo são ferramentas de busca, tais como as que permitem localizar imagens (a exemplo do Google), as quais já foram consideradas lícitas em determinadas circunstâncias e em certas jurisdições (como no caso *Perfect 10 v. Amazon*, visto no Capítulo 3).

Desta forma, ressalta-se a pertinência de uma discussão mais aprofundada acerca do grau de abrangência do termo “sistema de inteligência artificial” na referida seção do projeto, a fim de evitar um processo de sobreinclusão que afete procedimentos de mineração de textos e dados que não concorram com artistas e que possuam relevante interesse público.

5.4.2. Escopo do uso livre para os agentes de pesquisa

Observa-se que a previsão de um uso livre de obras para processos de coleta e análise manteve-se presente, em alguma medida, desde a primeira proposta legislativa sobre IA – manifestando, desde cedo, a necessidade de resguardo às atividades de pesquisa e conservação (Brasil, 2021c). No entanto, nota-se que o atual debate legislativo sobre este tópico se concentra sobre instituições de pesquisa, educação e conservação – não contemplando profissionais que não integrem tais instituições e estejam, portanto, trabalhando de forma autônoma.

Poder-se-ia argumentar que isto não implica necessariamente que tais indivíduos estejam proibidos de usar obras em pesquisas com IA – ao contrário, o art. 1º, §1º do PL nº 2.338/2023 determina que a regulação ali posta não se aplicaria a pesquisadores privados, o que também os excluiria, em tese, da proibição de uso de obras em IA e da obrigatoriedade de remuneração pelo mesmo.²⁴² Por outro lado, na ausência de uma previsão legal expressa que

²⁴² “Art. 1º

(...)

§ 1º. Esta lei não se aplica ao sistema de inteligência artificial:

a) usado por pessoa natural para fim exclusivamente particular e não econômico, salvo o disposto na Seção V do Capítulo IV - Medidas de Governança para Sistemas de Inteligência Artificial de Propósito Geral e Generativas - desta Lei; (...)

c) em atividades de testagem, desenvolvimento, pesquisa ou que não sejam colocadas em circulação no mercado, desde que mantida exclusivamente sua finalidade de investigação e desenvolvimento científico, sem prejuízo de observar a legislação pertinente, em especial a Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do

cubra estes profissionais, e considerando o art. 29 da LDA sobre a inclusão não-autorizada de obras em base de dados, isto significaria que aqueles que lidam com pesquisa e que não se encontram institucionalizados ainda se veriam em um cenário de incerteza quanto à regularidade de seu trabalho – cuja manutenção, como visto no Capítulo anterior, é importante para sua colocação profissional no âmbito das instituições cobertas pelo artigo em questão.

Ao mesmo tempo, observa-se a ausência, no debate legislativo, de referências à pesquisa jornalística, outrora contemplada no Substitutivo de 2022 (Brasil, 2022c). Como visto no Capítulo 1, este é um segmento que também tem se beneficiado de tecnologias de coleta e análise para diversos fins de relevância social, como o jornalismo de dados e a checagem de fatos – de forma que a sua reintrodução em um eventual texto legal se coloca como um ponto que, no entendimento deste trabalho, poderia gerar resultados positivos para o interesse público.

5.4.3. Questões de transparência

A exigência de que se divulgue o material protegido por Direito Autoral usado no treinamento de IA tem, a princípio, coerência prática: ajuda titulares a ter controle sobre o uso de sus obras e facilitaria a identificação de infrações. Também é consistente com os princípios gerais de transparência, explicabilidade e auditabilidade da IA, assim determinados pelo Projeto de Lei,²⁴³ bem como caminha lado a lado com o proposto em regulações como o *AI Act* europeu.

Contudo, conforme visto ao longo deste trabalho (vide Guadamuz, 2024 e Margoni, 2024), questiona-se se os meios atualmente disponíveis para identificar obras específicas incluídas em uma base de treinamento de inteligência artificial já teriam escopo e eficácia suficientes para cumprir seu papel a contento – especialmente na realidade dos megadados, em que milhões de materiais são o objeto de análises de dados e procurar obra por obra se torna, portanto, uma tarefa desafiadora.

No campo acadêmico, conforme observamos no Capítulo 4, isto se agrava pela falta da difusão de práticas de documentação da geração, obtenção e organização dos dados – atos estes que demandam investimento de tempo e esforço técnico (a ponto de, afinal, serem objeto

Consumidor) e a Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais) e a Lei nº 6.938, de 31 de agosto de 1981 (Política Nacional do Meio Ambiente)” (Brasil, 2024b).

²⁴³ Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:

(...)

VI - transparência e explicabilidade, observado o segredo comercial e industrial;

VII - diligência devida e auditabilidade ao longo de todo o ciclo de vida do sistema de inteligência artificial, de acordo com risco envolvido e o estado da arte do desenvolvimento tecnológico;

(...)

X - prestação de contas, responsabilização e reparação integral de danos;” (Brasil, 2024b).

específico de proteção na União Europeia, tal como visto no direito *sui generis* para bases de dados). Como vimos, trata-se de um problema duplo: não apenas os estudos analisados naquele capítulo apontam que este trabalho de documentação tem poucos aderentes, como também não parece haver um padrão único estabelecido para o tratamento de dados de pesquisa. Com efeito, criar um modelo deste tipo pode ser uma tarefa deveras desafiadora ante à diversidade de tipos de dados e de práticas metodológicas entre áreas do saber.

Há ainda desafios adicionais quando o material de análise são obras protegidas: Quintais menciona diversas características do próprio sistema de Direito Autoral que dificultam a aferição de transparência: o baixo grau de originalidade requerido para proteção por direitos autorais, “a fragmentação territorial do Direito Autoral e sua titularidade, a falta de uma exigência de registro para obras, e (...) a precariedade de metadados com informações sobre titularidade” (Quintais, 2023).

Isto causa certa preocupação quando se nota que a discussão legislativa sobre uma limitação para treinamento de IA por instituições de pesquisa e preservação da memória no Brasil não tem abordado a necessidade de resguardar estas entidades de tal obrigação – ao contrário do que se vê, por exemplo, no Artigo 53.1.c. do *AI Act* (União Europeia, 2024). Por exemplo, o artigo sobre transparência do PL nº 2.338/2023 não faz ressalvas explícitas nesse sentido, potencialmente reduzindo a eficácia da limitação para fins de pesquisa presente no Projeto. É pertinente, portanto, trazer esta questão de forma mais explícita na regulação, a fim de permitir uma melhor compreensão acerca dos direitos e obrigações envolvidos.

Por outro lado, também se nota que, caso a interpretação acima esteja correta, então também seria possível presumir que este dever de transparência se estende também a plataformas que licenciaram o uso de obras protegidas e também aos próprios titulares cessionários de direitos, caso financiem ou utilizem sistemas internos de inteligência artificial. Neste caso, observa-se que a discussão acima pode ter relevância não apenas para a pesquisa, mas também para questões relacionadas à exploração econômica das obras – as quais serão discutidas mais detalhadamente ao tratamos da remuneração, mais adiante.

5.4.4. Escopo dos atos permitidos

Em relação às finalidades dos usos a serem permitidos em relação ao treinamento de IA, nota-se um foco direcionado especificamente para o uso de obras protegidas “na pesquisa e desenvolvimento de sistemas de inteligência artificial” (Brasil, 2024b). Daí, é possível extrair duas interpretações: uma, mais extensiva, poderia levar à conclusão de que quaisquer atividades

de mineração de textos e dados na pesquisa estariam permitidas separadamente do treinamento por sistemas de IA. Já uma visão mais restritiva atrelaria essas atividades exclusivamente à pesquisa em IA por instituições de pesquisa e de preservação da memória – o que teria maior coerência com o fato de que os atuais projetos sobre o tema se destinam especificamente à inteligência artificial. Neste caso, regula-se apenas uma aplicação específica da mineração de textos e dados – que compõe apenas uma parcela dentro de um universo de atividades de pesquisa que utilizam estas ferramentas, mas não estão diretamente relacionadas à criação de IA.

Ademais, nota-se que não se pode apenas permitir o desenvolvimento de IA sem que também se permita pesquisa *com* IA, sob pena de deixar incerta a possibilidade de conduzir pesquisas envolvendo coleta e análise automatizada de dados, mesmo se o modelo a ser usado foi desenvolvido dentro da própria instituição. Nesse sentido, nota-se que o PL nº 2.338/2023 possuía um dispositivo específico nesse sentido na versão proposta pela Comissão no final de 2022, o qual não mais consta na versão mais recente: o art. 42, §2º previa a possibilidade de mineração de textos e dados automatizada de forma mais ampla, com fins de análise (Brasil, 2022c) – o mais próximo que se teve de uma permissão para mineração de textos e dados no Direito Autoral brasileiro, especialmente quando se considera o amplo escopo do que constitui inteligência artificial no atual debate regulatório brasileiro.

Por esta razão e pelas expostas acima, entende-se pertinente a discussão quanto a uma possível reintrodução deste dispositivo e uma melhor definição do escopo de atos permitidos pela limitação voltada para mineração de textos e dados em atividades de pesquisa, a fim de garantir sua eficácia e exequibilidade.

5.4.5. A exigência de finalidade não-comercial

Outra restrição no que diz respeito ao tipo de atividade permitida dentro do debate legislativo corrente é a exigência de que as atividades de pesquisa e desenvolvimento de IA pelas instituições contempladas tenham finalidade estritamente não-comercial, sem intuito de lucro. Tal proteção é, inclusive, corroborada pelo art. 65, que restringe expressamente o direito de remuneração ao treinamento de modelos de uso comercial. Segundo Flynn e Palmedo (2022), trata-se de uma exigência observada em alguns modelos regulatórios de limitações e exceções, inclusive entre países classificados como “verde” (o grau máximo de abertura à pesquisa) no estudo. Como já visto no Capítulo 3, a própria Diretiva CDSM, uma das inspirações mais aparentes do PL, também faz esta ressalva. A disposição tampouco configura elemento

estranho ao atual sistema brasileiro de limitações e exceções, que já possui hipóteses que excluem fins lucrativos.²⁴⁴

Não obstante, é importante destacar que atividades de pesquisa científica, mesmo em universidades e instituições similares, não operam estritamente dentro de um caráter não-comercial: conforme discutido no capítulo anterior, o paradigma de universidades e centros de pesquisa integrados ao ecossistema de inovação não configura uma tendência recente, com a própria noção de universidade empreendedora ocupando lugar cada vez mais garantido em regulações de incentivo à inovação. O Brasil, como observado, também tem se movimentado nas últimas décadas no sentido de estimular o diálogo com empresas e a criação de empresas spin-off a partir das universidades, com a propriedade intelectual figurando como item de destaque. Ademais, Ascensão (2002) lembra que mesmo a simples publicação de um trabalho científico, muitas vezes, envolve atividades comerciais, por si só, como periódicos fechados ou mesmo a venda de livros.

A presença desta exigência, portanto, levanta diversas questões que necessitam ser enfrentadas: como isto afeta o trabalho de escritórios de transferência de tecnologia e o processo de regularização empresarial de um projeto que lida com inteligência artificial em uma universidade? E como esta exigência afeta as vias possíveis de divulgação científica? Seria possível publicar a respeito do desenvolvimento de um sistema de IA somente em periódicos abertos, mesmo com a exigência de priorizar periódicos fechados, de maior fator de impacto?

Desta maneira, entende-se que este tipo de disposição se beneficiaria de um debate mais aprofundado, considerando o possível impacto dessas questões para as políticas de inovação assumidas pelo Estado brasileiro e o escopo já elevado de restritividade às hipóteses permitidas pelo artigo em discussão.

5.4.6. O uso da regra dos três passos

Outra questão é vista na previsão explícita de elementos da regra dos três passos dentro das propostas regulatórias mais recentes sobre IA, o que confere a um uso livre fechado os critérios para legalidade típicos de uma cláusula aberta. Isto traz um problema de aplicação da norma: de uma forma geral, cláusulas abertas, tal como visto no Capítulo 3, geralmente seguem

²⁴⁴ Ver, por exemplo, o art. 46 da LDA: “Não constitui ofensa aos direitos autorais: (...) II - a reprodução, em um só exemplar de pequenos trechos, para uso privado do copista, desde que feita por este, sem intuito de lucro; (...) VI - a representação teatral e a execução musical, quando realizadas no recesso familiar ou, para fins exclusivamente didáticos, nos estabelecimentos de ensino, não havendo em qualquer caso intuito de lucro;” (Brasil, 1998b).

critérios amplos, principiológicos, que servem de baliza para a criação de hipóteses legais mais específicas (seja pelo Legislativo ou pelo Judiciário) em um momento posterior. Na prática, sugere-se, assim, que a conformidade da mineração de textos e dados para o treinamento de IA por instituições de pesquisa com a exploração normal da obra ou com os interesses legítimos do titular não é certa, mas dependerá do caso concreto. Em outras palavras, deverá ser definida por terceiros, seja por contrato ou pela via judicial. Note-se que a própria União Europeia (2019) já omitia a menção à regra de sua limitação para pesquisa, e a legislação japonesa, ainda que faça alusão à regra, qualifica-a dentro de uma interpretação voltada para o uso pretendido da obra original (Japão, 2018).

Isto é um problema quando se considera que o estabelecimento do que conta como “exploração normal da obra” é, ao fim e ao cabo, uma questão a ser definida por práticas de mercado. Nesse sentido, se o licenciamento de bases de dados ou obras para fins de uso em mineração de textos e dados ou *machine learning* se tornar prática de mercado corriqueira, isto poderia configurar uma “exploração normal da obra” – e, portanto, fora do escopo da limitação, mesmo se o uso não implicar em qualquer efeito disruptivo ou concorrencial para o artista ou o titular. É importante lembrar que, como visto no Capítulo 1, já existem plataformas (como a WIPOLex) cujos termos de uso restringem o uso de obras no acervo para estes fins.

Neste caso, convém discutir alternativas redacionais mais precisas, que evitem a possibilidade de esvaziamento da limitação em decorrência de mudanças no mercado. Uma possível solução seria substituir “exploração normal da obra” por uma definição que exclua apenas usos generativos ou expressivos: por exemplo, “exploração normal da expressão contida nas obras”.

Ao mesmo tempo, outra observação mais específica se faz relevante: nota-se que, diferentemente do PL nº 21/2020 (Brasil, 2021a), a versão mais recente do PL nº 2.338/2023 (art. 63, IV) faz referência mais completa à regra dos três passos. No entanto, ali se prescreve que o uso livre previsto para pesquisa não deve prejudicar injustificadamente os interesses *econômicos* dos titulares (Brasil, 2024c), e não os interesses *legítimos*, como no texto da Convenção de Berna (explorado no Capítulo 2). Considerando-se que os interesses ali resguardados não se limitam necessariamente a questões patrimoniais, cumpre reanalisar a redação do artigo também para evitar uma invasão indevida da limitação aos interesses legítimos dos autores, sobretudo em relação aos interesses de ordem não-patrimonial.

5.4.7. Questões de clareza sobre a exigência de acesso lícito

Outra grande questão suscitada pelo atual debate legislativo no Brasil é a presença da exigência de acesso lícito, de maneira similar ao observado no Art. 3 da Diretiva CDSM. Reiteramos, aqui, as considerações feitas àquele dispositivo no Capítulo 3, especialmente as críticas ali postas em relação à falta de clareza legal quanto ao modo exato de aferir essa licitude. No caso brasileiro, isto é agravado pelo fato de que não se observa, atualmente, nenhuma definição legal do que configuraria acesso lícito nesse contexto. Com efeito, tal como escrito, a aplicação dos dispositivos atualmente propostos no âmbito do Legislativo dependeria da definição apresentada na União Europeia – a qual não apenas mantém as instituições reféns de contratos restritivos ou termos de uso ambíguos (Schirru e Margoni, 2023) como deposita sobre o usuário da limitação a tarefa hercúlea, para não dizer impossível, de verificar o status legal de cada obra incluída na análise (Margoni, 2024).

Aqui, cabe lembrar que, nas entrevistas utilizadas neste trabalho, os respondentes, em sua maioria, declaram já enfrentar problemas relativos a obstáculos institucionais e contratuais para o acesso a trabalhos e dados científicos, à falta de liberdade para compartilhar acervos e, principalmente, à incerteza em relação à legalidade das suas atividades na ausência de regras mais claras na lei. Nesse sentido, portanto, entende-se que o requisito de “acesso lícito”, tal como proposto, se encontra em uma capacidade reduzida para solucionar alguns dos principais obstáculos para a pesquisa científica apontados pelas entrevistas.

5.4.8. A possibilidade de sobreposição contratual

Para além de questões conceituais, um desdobramento da exigência de acesso lícito, também discutido no Capítulo 3, é a sua possibilidade de servir, na prática, como uma imposição explícita de observância de eventuais arranjos privados para o acesso e uso de um determinado material – efetivamente, colocando as L&E em papel de subserviência às demais camadas de proteção descritas por Dusollier (2005) e Derclaye (2008), sendo uma delas os arranjos contratuais.

Neste sentido, o fato de a definição de “acesso lícito” da Diretiva CDSM estipular o licenciamento do conteúdo como uma hipótese de satisfação dessa condição (nos casos em que a obra não foi tornada publicamente disponível) leva a uma ponderação: ainda que se estabeleça que o uso livre para mineração de textos e dados ou para treinamento de inteligência artificial seja imune à reserva legal garantida ao titular – o que, em tese, tornaria nulas quaisquer disposições contratuais proibindo a mineração para o desenvolvimento de IA –, a exigência de acesso

lícito mantém uma possibilidade indireta de afastamento da limitação pela via contratual. Trata-se de algo inédito no regime brasileiro de limitações e exceções, fortalecendo a noção de controle sobre o acesso como prerrogativa do titular, ainda que não seja um direito exclusivo reconhecido em lei (Derclaye, 2008).

No presente trabalho, entende-se que, em circunstâncias normais, a possibilidade de sobrepor uma hipótese de uso livre na LDA pela via contratual seria incompatível com o respeito à função social da propriedade prevista na Constituição, bem como fugiria da observância à função civil dos contratos e à exigência de licitude do objeto do negócio jurídico, conforme previsto no Código Civil brasileiro.²⁴⁵ Contudo, a imposição explícita de acesso lícito, até então ausente de outras limitações e exceções no Direito Autoral pátrio, abriria a possibilidade de uso desse tipo de disposição – o que se torna problemático quando se consideram as relações assimétricas com grandes titulares e repositórios estrangeiros relatadas nas entrevistas analisadas no Capítulo anterior. Este cenário se agrava com a ausência, na lei autoralista brasileira, de uma disposição explícita que proíba a sobreposição contratual das L&E – na contramão de legislações promulgadas em dezenas de países (Band, 2023).

Desta maneira, entende-se relevante uma discussão mais aprofundada sobre a presença e aplicação da exigência de acesso lícito, uma vez que o dispositivo, tal como posto, pode reduzir significativamente a eficácia da hipótese de uso lícito prevista no artigo e, ademais, estabelecer um precedente jurídico que restrinja o exercício das limitações e exceções já consagradas na legislação.

5.4.9. Medidas tecnológicas de proteção

Outro desdobramento da cláusula de acesso lícito é o reforço não apenas da abertura do uso livre para pesquisa à sobreposição contratual, mas também de outro elemento problemático: o fato de a LDA (art. 107, já visto anteriormente) não prever qualquer possibilidade de contornar mecanismos de proteção tecnológica, mesmo para o propósito de lançar mão das limitações e exceções dos arts. 46 a 48.

Em um contexto de IA Generativa, observa-se que já há ferramentas destinadas a dificultar a captura não-autorizada de obras protegidas por ferramentas de coleta e análise de dados. Nas artes visuais, um exemplo é o Glaze Project,²⁴⁶ que desenvolve ferramentas como Glaze e

²⁴⁵ “Art. 104. A validade do negócio jurídico requer: (...) II - objeto lícito, possível, determinado ou determinável; (...)

Art. 421. A liberdade contratual será exercida nos limites da função social do contrato.” (Brasil, 2002).

²⁴⁶ Disponível em: <https://glaze.cs.uchicago.edu/aboutus.html>

Nightshade: softwares que podem ser utilizados para modificar imagens de forma que estas permaneçam as mesmas a olho nu, mas contenham pequenos elementos em sua composição que alteram suas características quando são copiadas e processadas por ferramentas de *scraping*. Segundo os desenvolvedores (ao descrever o Nightshade, mas mencionando também o Glaze), estes programas buscam se dirigir a uma assimetria de forças na qual artistas têm seu trabalho utilizado sem consentimento, listas de opt-out (como a prevista na União Europeia) são desrespeitadas e a falta de transparência impede a identificação de violações. A intenção é elevar os custos de uso não-licenciado – tornando o licenciamento, portanto, “uma alternativa viável”.²⁴⁷²⁴⁸ Perceba-se que não há, no site do projeto, qualquer menção a limitações e exceções.

Cabe ressaltar que tais ferramentas, ao menos no caso do Shade Project, foram criadas para contextos como o norte-americano, onde não há restrições tão severas para o uso de obras – não há nenhuma disposição sobre inclusão de obras em bases de dados como direito exclusivo, e a licitude da mineração de textos e dados, conforme já visto, dependerá do caso concreto, com certos casos cobertos pela doutrina do uso justo. Ademais, tem-se o fato de, nos EUA, o *fair use* não se estender ao DMCA (que ainda conta com uma lista fechada de limitações), além do já mencionado custo financeiro envolvido em definir o escopo de limitações pela via judicial. Neste cenário, faz sentido que artistas recorram a tal tipo de ferramenta como mecanismo preventivo para coibir usos potencialmente danosos de suas obras – e, assim como processos de mineração teriam dificuldade de averiguar a legalidade de cada obra incluída em condições usuais de pesquisa, o uso desses mecanismos de proteção ainda não parece permitir uma distinção precisa de quais usos seriam legítimos ou não.

Aqui, observa-se um dilema. Por um lado, o ambiente regulatório descrito acima é específico de um país (os EUA), mas difere muito de outros locais: por exemplo, na União Europeia (que, por si só, já engloba múltiplas jurisdições nacionais, cada uma com sua

²⁴⁷ “Since their arrival, generative AI models and their trainers have demonstrated their ability to download any online content for model training. For content owners and creators, few tools can prevent their content from being fed into a generative AI model against their will. Opt-out lists have been disregarded by model trainers in the past, and can be easily ignored with zero consequences. They are unverifiable and unenforceable, and those who violate opt-out lists and do-not-scrape directives can not be identified with high confidence.

(...)

Used responsibly, Nightshade can help deter model trainers who disregard copyrights, opt-out lists, and do-not-scrape/robots.txt directives. It does not rely on the kindness of model trainers, but instead associates a small incremental price on each piece of data scraped and trained without authorization. Nightshade's goal is not to break models, but to increase the cost of training on unlicensed data, such that licensing images from their creators becomes a viable alternative.

²⁴⁸ Disponível em: <https://nightshade.cs.uchicago.edu/whatis.html>

regulação específica), opera-se com uma lista fechada de limitações e exceções, mas a Diretiva de 2019, conforme visto, torna explícito que quaisquer ferramentas de proteção tecnológica devem permitir o livre exercício das L&E ali previstas. E o *AI Act* prevê uma cláusula exigindo transparência em relação ao material usado para o treinamento de inteligências artificiais.

No Brasil, como visto ao longo deste trabalho, tem-se um regime muito distinto dos EUA: no momento, atividades de coleta e análise de dados que envolvam obras protegidas são efetivamente proibidas, não há qualquer uso livre previsto para atividades de pesquisa na legislação em vigor e a atual proposta em discussão no Congresso Nacional também prevê a transparência do material usado, além de restringir a limitação proposta ao uso em pesquisa para treinamento de IA por entidades sem fins lucrativos – embora seja possível reconhecer que mesmo a categoria mais aberta de L&E para mineração de textos e dados proposta por Flynn et al (2023) comportaria a proibição de uso comercial. Outra especificidade brasileira é que nosso correspondente nacional para o DMCA – o art. 107 da LDA – não é aberto para as limitações e exceções hoje previstas em lei (tal como ocorre no Reino Unido), tampouco possui uma relação de usos livres própria (como nos EUA).

Por outro lado, o ambiente sobre o qual estas regulações operam não é restrito aos seus territórios: a internet é compartilhada por usuários de todo o mundo, e tanto os processos de coleta e análise quanto o desenvolvimento e operação de sistemas de IA podem se dar de (e em) diversas partes do globo. É razoável presumir que artistas de países externos aos EUA também queiram se resguardar de usos disruptivos empreendidos por empresas estadunidenses – e não se vê razão para que não se incluam brasileiros entre estes criadores.

O resultado desta lógica, no caso brasileiro, é o uso de ferramentas criadas para um ambiente jurídico mais permissivo, porém em um contexto regulatório muito mais restritivo para instituições de pesquisa, jornalistas e entidades de conservação da memória. Nesse cenário, portanto, evidencia-se novamente a importância de uma ponderação maior acerca da adequação das medidas anti-burla hoje existentes no Direito Autoral brasileiro ao exercício das L&E, sob pena de esvaziar sua eficácia– e, com isso, os direitos dos usuários e o interesse público.

5.4.10. A natureza e aplicação da reserva legal

Mais recentemente, observa-se que a discussão sobre a criação de um instrumento legal que permite ao titular proibir o uso de obras protegidas para treinar inteligências artificiais tem assumido, essencialmente, a forma de uma reserva legal opcional, similar ao previsto no art. 4(3) da Diretiva CDSM a respeito da mineração de textos e dados.

Aqui, nota-se a pertinência de se incluir, neste debate, uma definição legal a respeito de como esta proibição deve ser implementada para gerar efeitos jurídicos: se basta um anúncio informal (por exemplo, em perfil de rede social), se é necessário estipulá-la contratualmente ou se é preciso estabelecê-la através de um formato legível por máquinas, por exemplo. Do contrário, somam-se mais incertezas a um critério que, como discutido a respeito da regulação europeia, já apresenta dificuldades inerentes à sua concretização, especialmente no que diz respeito à sua possibilidade de retroagir para modelos já treinados, à verificação da licitude de duplicatas no treinamento (Keller, 2024a; 2024b). Por outro lado, esta omissão também representaria um obstáculo ao titular de direitos, uma vez que não haveria como saber, com certeza, se a reserva legal foi feita de forma legalmente válida.

Outra questão relacionada é o alcance de atos afetados pela reserva legal: diferentemente do art. 29 da LDA, que requer do usuário a prévia e expressa autorização do autor para os usos ali elencados (um *direito de autorizar*), este tipo de norma, nos moldes atuais (por exemplo, ver Brasil, 2024c), confere um *direito de proibir*. Em outras palavras, é o detentor de direitos quem precisa se manifestar contrário ao uso de sua obra em IA para que seu interesse seja juridicamente validado – caso contrário, o uso da obra seria lícito.

Ao mesmo tempo, isto leva a outra indagação: até que ponto seria possível alcançar usos ocorridos antes de tal manifestação pelo titular? Nesse sentido, nota-se que o PL nº 2.338/2023 prevê, atualmente, a retroatividade da responsabilidade penal por perdas e danos resultantes do treinamento mesmo antes da proibição ser efetivamente publicada. Contudo, considerações feitas acima sobre a cláusula de transparência, bem como os dilemas envolvendo a comprovação de dados nos casos judiciais descritos no Capítulo 1, sugerem que demonstrar tais prejuízos não seja tarefa simples no momento. Ainda assim, por esta lógica, pareceria mais seguro considerar que quaisquer obras fora do domínio público necessitariam de autorização prévia e expressa pra uso, como nos direitos expostos na LDA.

Estes problemas, somados à já observada variedade de modelos e cláusulas contratuais existentes no mercado (Schirru e Margoni, 2023) agravam a incerteza regulatória criticada nas entrevistas. Faz-se, portanto, necessária uma revisão e elucidação deste e de outros pontos, a fim de facilitar a compreensão adequada dos direitos e obrigações ali implicados e, ao mesmo tempo, viabilizar o seu cumprimento e trazer maior segurança não apenas para os titulares, mas também para os agentes de pesquisa.

5.4.11. Escopo do dever de remuneração

Se o debate sobre a implementação de usos livres para fins de pesquisa já suscita certas questões em relação ao escopo do que seria ali permitido, também é necessário definir o escopo de quais atos envolvendo mineração de textos e dados – e inteligência artificial – ensejam a necessidade de remuneração. No caso específico do PL nº 2.338/2023, tais atos englobam “processos de mineração, treinamento ou desenvolvimento de sistemas de IA” (Brasil, 2024c), sem deixar claro se a menção à mineração deve se referir estritamente ao uso desta para o desenvolvimento de inteligência artificial ou se diria respeito a qualquer processo de análise.

Esta diferenciação possui significativos impactos regulatórios: se a segunda interpretação acima for adotada, o dever de remuneração teria, então, o efeito de impor um dever de remuneração a instituições de pesquisa, jornalismo ou conservação que fizerem uso de análise automatizada que não envolva o desenvolvimento de IA – o que poderia representar um retrocesso em relação ao atual regime jurídico e reforça a necessidade de disposições que cubram a mineração de textos e dados para além do simples desenvolvimento ou treinamento, tal como discutido no item 5.4.4. Por outro lado, seguindo-se uma interpretação mais restritiva, a aplicação mais correta deste dispositivo parece ser aquela que estende a exigência deste dever de remuneração apenas para mineração no processo de desenvolvimento de inteligências artificiais. De qualquer forma, reitera-se a necessidade de uma redefinição dos atos que demandariam remuneração, a fim de garantir o efetivo exercício de qualquer uso livre a ser previsto para a mineração de textos e dados em atividades de pesquisa.

5.4.12. Novos direitos exclusivos: quem é o titular?

Independentemente das questões impostas pela reserva legal e pelo dever de remuneração, contudo, é possível extrair uma conclusão: tal como postos no atual debate regulatório, os dispositivos atualmente em discussão parecem conferir ao detentor de direitos uma prerrogativa sobre o uso da obra em IA que existe independentemente de haver autorização, por exemplo, para o acesso e inclusão de uma obra em base de dados. Na prática, portanto, entende-se possível a interpretação de que se esteja diante de novos direitos exclusivos, distintos dos demais direitos já previstos em lei aos autores e outros titulares.

Esta observação ganha contornos problemáticos quando se nota que, diferente do art. 29 da LDA, que afirma que qualquer ato ali elencado “depende de autorização prévia ou expressa *do autor*” (Brasil, 1998b, destaque nosso), propõe-se, atualmente, que o exercício desse novo direito caberia *ao titular*. Considerado de forma isolada, pode-se entender que este direito

seria garantido a qualquer um que detenha a titularidade sobre os direitos econômicos – e esta, embora originariamente caiba ao autor, muitas vezes, está concentrada nas mãos de editoras, gravadoras ou outros tipos de empresa, por meio de contratos de cessão de direitos. Nesse sentido, o PL corre o risco de preterir diversos artistas, que sentem mais diretamente os efeitos da IA Generativa, em favor de grandes titulares (Senftleben, 2024a).

No entanto, cabe manifestar um entendimento diverso: se a reserva legal e o dever de remuneração forem, de fato, novos direitos exclusivos, a presunção deverá ser de que, a menos que expressamente previsto em licença ou cessão de direitos, esta reserva legal cabe ao titular originário da obra: o autor, nos termos do art. 24 da Lei nº 9.610/98. Isto porque a LDA considera cada modalidade de uso da obra independente das demais, devendo constar de forma explícita e individualizada na licença ou cessão;²⁴⁹ da mesma forma, proíbe-se expressamente a transferência automática de modalidades de uso não previstas à época do contrato,²⁵⁰ bem como exige-se que os negócios jurídicos envolvendo direitos autorais sejam interpretados de forma restritiva.²⁵¹ Em outras palavras, estamos diante de uma prerrogativa ainda não cedida ou licenciada, e que, por força de lei, ainda pertenceria aos autores – tornando o uso da palavra *titular*, neste contexto, inadequado ao ordenamento jurídico vigente.

Adicionalmente, o uso do termo *titular*, caso seja interpretado para ser aplicado a quem quer que possua os direitos patrimoniais sobre uma obra naquele momento, pode gerar uma implicação preocupante: a de que o autor, caso ceda estes direitos a um terceiro, não poderá se opor ao uso de sua obra por IA por este mesmo terceiro no futuro. Isto pode colocar artistas, pesquisadores e outros criadores em posição vulnerável diante dos efeitos potencialmente substitutivos da IA Generativa caso esta seja empregada pelo próprio titular de direitos – a menos que haja uma cláusula legal prevendo, por exemplo, a intransferibilidade desse direito. É importante recordar que o uso de sistemas de IA por grandes titulares foi uma questão levantada na greve dos atores e roteiristas nos EUA, conforme visto no Capítulo 1, e que diversas editoras acadêmicas detêm direitos patrimoniais sobre diversos trabalhos científicos e vêm desenvolvendo seus próprios programas de IA, como observado no Capítulo 4.

²⁴⁹ “Art. 31. As diversas modalidades de utilização de obras literárias, artísticas ou científicas ou de fonogramas são independentes entre si, e a autorização concedida pelo autor, ou pelo produtor, respectivamente, não se estende a quaisquer das demais.” (Brasil, 1998b).

²⁵⁰ “Art. 49. Os direitos de autor poderão ser total ou parcialmente transferidos a terceiros, por ele ou por seus sucessores, a título universal ou singular, pessoalmente ou por meio de representantes com poderes especiais, por meio de licenciamento, concessão, cessão ou por outros meios admitidos em Direito, obedecidas as seguintes limitações: (...)”

V - a cessão só se operará para modalidades de utilização já existentes à data do contrato;” (*Ibid.*).

²⁵¹ “Art. 4º Interpretam-se restritivamente os negócios jurídicos sobre os direitos autorais.” (*Ibid.*)

Entende-se, portanto, que o ajuste no uso deste termo facilitaria a compreensão quanto a quem detém a prerrogativa de controlar o uso de uma obra para o treinamento de IA e ajudaria a trazer maior resguardo jurídico aos autores neste processo.

5.4.13. O licenciamento como alternativa – como estipular a remuneração?

Com o desenvolvimento e difusão de medidas tecnológicas de proteção para barrar a coleta de obras para mineração de textos e dados e treinamento de IA, bem como a intransponibilidade dos dispositivos anti-burla na LDA, o licenciamento para o uso de obras nesses casos se coloca como uma opção interessante. Assim, voltamos nossa atenção para o art. 65, que trata especificamente do dever de remuneração pelo uso de obras protegidas no treinamento de IA, para discutir de que forma este processo poderia ser viabilizado – especialmente reconhecendo a dificuldade técnica em identificar, localizar e negociar com todos os titulares das obras envolvidas, conforme discutido anteriormente neste trabalho.

Para além da indagação de quem deve ser remunerado – que toca diretamente a discussão sobre *autores e titulares* comentada acima –, uma questão crucial será definir, em primeiro lugar, qual o modelo a ser adotado para administrar o licenciamento, a remuneração, a arrecadação e repasse dos valores devidos aos titulares. Uma opção seria o estabelecimento de fundos de compensação sob responsabilidade de entidades de gestão coletiva de direitos autorais, ou associações de titulares de direitos de autor e conexos, cujo funcionamento é previsto nos artigos 97 a 100-B da Lei no. 9.610/98. Tais entidades teriam a vantagem de já possuir legitimidade legal, regulamentação e recursos para realizar este trabalho. Por outro lado, resta definir até que ponto é possível garantir que a via da gestão coletiva seja capaz de criar sistemas de remuneração efetivos, claros e transparentes – especialmente no Brasil, que não é estranho a questionamentos envolvendo a transparência e lisura no processo de arrecadação e distribuição via gestão coletiva.²⁵²

Outra alternativa é considerar a possibilidade de negociação direta com agentes de IA, o que inclusive é previsto pelo art. 65, inciso I do PL nº 2.338/2023 ao resguardar a liberdade dos titulares “de negociar e autorizar diretamente ou coletivamente” o uso de suas obras pelos agentes de IA, “independentemente de regulamentação posterior” (art. 65, §2º). No entanto, é necessário lembrar, mais uma vez, da posição desvantajosa enfrentada pelos artistas – na

²⁵² Um exemplo foi a realização de Comissão Parlamentar de Inquérito (CPI), em 2012, diante de supostas irregularidades cometidas pelo Escritório Central de Arrecadação e Distribuição (ECAD). Para mais informações, ver: <https://www.camara.leg.br/noticias/372144-relatorio-final-da-cpi-do-ecad-chega-a-camara/>. Acesso em 08 out. 2024.

maioria, desprovidos dos meios econômicos e jurídicos acessíveis a grandes plataformas e empresas titulares de direitos (Senftleben, 2024).

Outro elemento a ser considerado ao se discutir modelos de licenciamento é, naturalmente, como efetuar a remuneração pelo uso em sistemas de IA. Alguns dos principais termos a serem considerados para tanto estão elencados de forma exemplificativa no art. 64, §1º, II do PL nº 2.338/2023: o porte do agente de IA e os efeitos concorrenciais dos resultados em relação aos conteúdos originais utilizados.

Nesse sentido, a averiguação do efeito concorrencial pode se mostrar desafiadora quando consideramos as complexidades envolvidas no cumprimento do dever de transparência: como averiguar os danos causados se a simples identificação de quais obras foram usadas não costuma ser uma tarefa simples? Afinal, é importante lembrar que inteligências artificiais necessitam de uma quantidade humanamente incomensurável de dados para funcionar adequadamente, e os modelos finais não retêm as obras de fato, e sim um sistema de pesos que reduz diversas obras a uma “massa” praticamente indistinta (Guadamuz, 2024).

Ademais, ainda que fosse possível identificar o uso obra por obra, há outro problema: como calcular a remuneração? Afinal, segundo observa Guadamuz (2023), um livro com 72.000 palavras, ou 96.000 tokens, corresponderia a 0,00000007% da totalidade de um modelo como o GPT-4, que possui mais de 13 trilhões de tokens (OpenAI, 2024d).²⁵³ Ainda que este livro seja usado com grande frequência, é perfeitamente lógico deduzir que ao autor caberá, na melhor das hipóteses, um valor de *centavos* pela inclusão da obra no modelo. Nestes termos, seria mais fácil empregar o dever de remuneração como mecanismo de extinção do agente de IA do que como uma ferramenta de remuneração apropriada aos artistas.

Por outro lado, é importante considerar que mesmo um agente de baixo poder econômico (por exemplo, uma *startup* originada de um grupo de estudantes incubados em uma universidade) pode lidar com uma quantidade enorme de dados e, ao mesmo tempo, deter uma capacidade financeira muito reduzida de pagar pelo uso de obras. Assim, mesmo que se respeite o princípio da proporcionalidade (como sugere a consideração e se cobre um valor reduzido pelo uso, está-se diante de um cenário que prejudica a todos os envolvidos: a falência de um negócio a troco de pouco ou nenhum retorno financeiro aos titulares.

Também se observa que, na ausência de ressalvas de gratuidade a IAs não-generativas ou que não façam uso expressivo das obras, qualquer desenvolvimento de IA com aplicação

²⁵³ Ver: <https://www.gptcostcalculator.com/open-ai-token-calculator>. Acesso em 03 out. 2024.

comercial que inclua obras protegidas ensejará remuneração. Trata-se, portanto, de uma cobrança que, analisada em conjunto com o amplo escopo do que é considerado inteligência artificial nos termos regulatórios mais recentes, potencialmente estende o dever de remuneração para além do atendimento dos interesses feridos pela IA Generativa, impondo obstáculos desnecessários sobre atividades que não têm qualquer efeito substitutivo sobre os artistas.

Desta forma, faz-se necessário pensar e discutir alternativas ao pagamento que possam servir mais adequadamente aos interesses dos artistas. Isto pode envolver um modelo misto, mesclando fundos de gestão coletiva e negociação direta, ou mesmo incluir sistemas distintos: por exemplo, pela via tributária (Senftleben, 2024) ou mesmo pela criação de uma renda básica. De qualquer maneira, observa-se a necessidade de se discutir em maior detalhe não apenas a imposição do dever de remunerar, mas a criação de regras claras e justas sobre *como* remunerar.

5.4.14. A posição do Estado como mediador

Um ponto mais específico do atual debate regulatório sobre IA diz respeito ao próprio papel do Estado como intermediador no processo de governança dessas questões. No PL nº 2.338/2023, isto se exemplifica com a questão do *sandbox* regulatório, que serviria como plataforma de experimentos estatais para estabelecer regras de governança da IA. Conforme mencionado anteriormente, isto se estendia à possibilidade de intervenção nos critérios a serem seguidos para o cumprimento das obrigações de transparência e de remuneração pelo uso de obras protegidas no treinamento de inteligência artificial.²⁵⁴

Entretanto, isto foi suprimido da versão aprovada pelo Senado, sobrevivendo apenas uma estipulação de que os critérios de transparência ficarão a cargo de um “regulamento específico”. Isto pode ser interpretado como uma transferência das prerrogativas sobre remuneração aos agentes privados. Isto parece ressoar com uma preocupação em assegurar a capacidade de livre negociação entre as partes – o que é explicitamente mencionado no próprio art. 65, III da versão aprovada pelo Senado do mesmo Projeto de Lei.

Isto implica em dizer que diversos termos jurídicos com definição incompleta ou problemática presentes ao longo da seção sobre direitos autorais, bem como as diversas questões acerca de quem pagará pelo uso, quem receberá e como a remuneração será gerida, também

²⁵⁴ “Art. 64. A autoridade setorial, ouvido o órgão central competente, estabelecerá um ambiente regulatório experimental (*sandbox*) para dispor sobre a transparência e a remuneração devida por agentes de sistemas de inteligência artificial em relação a conteúdos protegidos por direitos autorais utilizados e treinados no desenvolvimento de sistemas de inteligência artificial disponibilizados com finalidade comercial, observado o disposto nesta Seção.” (Brasil, 2024e. Destaque feito pelo autor). O artigo prossegue elencando as mesmas diretrizes observadas no atual art. 64, observadas anteriormente.

ficarão a cargo dos titulares e dos agentes de IA. Em nome da autonomia entre as partes e da liberdade contratual, portanto, corre-se o risco de aprofundar ainda mais o cenário de incerteza regulatória já acusado em outros pontos. No caso específico do PL nº 2.338, isto é agravado pela imposição de que a validade dos negócios jurídicos oriundos destas relações se manteria mesmo perante regulações futuras, retirando do Estado a possibilidade de efetuar correções regulatórias neste cenário (art. 65, §2º) (Brasil, 2024c).

Isto causa preocupação quando se tem em mente, mais uma vez, a disparidade financeira e jurídica em que se encontram muitos artistas e usuários de pequeno porte, que estarão sujeitos a cláusulas desvantajosas impostas por grandes titulares e plataformas, beneficiadas pela falta de clareza regulatória. Mesmo em um cenário em que o Poder Judiciário seja levado a dirimir algumas dessas questões, é importante considerar – como visto nas observações sobre o modelo de cláusula aberta no Capítulo 3 – que a disparidade econômica também influencia a disposição e capacidade para se enfrentar um litígio. A retirada de opções regulatórias nesse sentido, portanto, pode dificultar a possibilidade de o Estado estabelecer garantias de interesse do interesse público – incluindo a justa remuneração de autores – sobre a exploração de obras protegidas no âmbito da inteligência artificial.

De uma forma mais ampla, observa-se um esforço em atender aos anseios dos titulares de direitos de autor e conexos em resposta aos desafios da IA Generativa, sem, no entanto, considerar as implicações práticas relativas à viabilidade técnica do cumprimento das obrigações legais a serem impostas, o estabelecimento de mecanismos legais que garantam a isonomia entre autores e cessionários de direitos e, por fim, o resguardo devido a atividades de interesse público que não trazem prejuízo aos interesses dos autores. E, no cerne deste problema, está a persistência de um modo particular de se pensar o Direito Autoral que entra em conflito com a realidade jurídica e material na qual se encontra.

5.5. Caindo nas mesmas armadilhas?

A discussão regulatória observada no Brasil em relação ao Direito Autoral ao longo deste Capítulo – e ao papel da mineração de textos e dados nesse cenário – revela uma tendência também notada em outras regulações analisadas neste trabalho: enquanto a década passada testemunhou movimentos no sentido de reajustar o sistema de Direito Autoral para um contexto de incentivo à pesquisa e à inovação, os últimos anos viram uma retração deste processo, motivada por preocupações com os efeitos concorrenciais da inteligência artificial, que tem resultado em uma pressão renovada por uma maximização dos direitos autorais. Ao mesmo tempo,

observa-se, no instituto, a continuidade de um modelo tradicionalista de propriedade, que coloca o interesse público e a participação de outros direitos fundamentais em termos nebulosos.

O debate em torno do Projeto de Lei nº 2.338/2023 se torna um exemplo dessa tensão, de forma implícita ou explícita, de diversas maneiras: por um lado, sob o pretexto de remoção de barreiras regulatórias à inovação, diversas propostas de emenda ao PL vão no sentido de apagar total ou parcialmente os dispositivos sobre direitos autorais, em um movimento de desregulação que simplesmente concederia a grandes empresas de tecnologia uma ampla liberdade sobre a condução da gestão dos dados e dos materiais utilizados na construção dos seus ativos. É importante reconhecer que a adoção desta postura em nada beneficia os agentes de pesquisa, que permaneceriam sem garantias legais para suas atividades, e deixariam autores vulneráveis a diversos usos não-autorizados de suas obras que diretamente concorreriam com o seu sustento, prejudicando o incentivo à produção de conhecimento e cultura.

Por outro lado, na tentativa de se colocar de forma diametralmente oposta a esse movimento, observa-se, particularmente, que a postura do Estado dentro da discussão sobre a mineração de textos e dados e inteligência artificial na esfera legislativa pátria parece ter assumido uma lógica mais tradicionalista do Direito Autoral, que termina por elevá-lo, na prática, a um *status* de direito universal, pessoal, natural e absoluto – desconectado do contexto histórico sobre o qual atua, da realidade material dos agentes que pretende regular e do regime jurídico no qual está inserida (Lewicki, 2007). Esta postura pode ser vista em diversos momentos, tais como na inclusão do acesso lícito (até então inexistente em qualquer outra limitação ao Direito Autoral pátrio) como requisito para a limitação do art. 63, na extensão da hipótese do mesmo artigo ao desenvolvimento específico de sistemas de IA dedicados a auxiliar no combate a ilícitos relacionados especificamente a direitos de autor e, mais indiretamente, na exclusão da ressalva legal sobre a LGPD no substitutivo de 24 de abril de 2024, que retornou nas versões seguintes do Projeto.

Ao mesmo tempo, uma nota técnica publicada recentemente pelo Ministério da Cultura, referente à regulação do uso de dados pessoais no treinamento de IA (assunto este distinto de questões autorais), parece confirmar a presença deste posicionamento no âmbito estatal ao entender que o estabelecimento de novos usos livres pelo Judiciário estaria sujeito somente à regra dos três passos e à efetivação de um direito fundamental – ignorando diretamente que o Tribunal deve observância aos direitos fundamentais mas também à *função social da propriedade*, conforme inclusive citado pelo próprio Ministério:

4.2.18 (...) No entanto, entendimento consolidado do **Superior Tribunal de Justiça admite novas limitações dos direitos exclusivos de autor para além das dispostas em lei, desde que atenda aos requisitos do “teste dos três passos” e o uso tenha como finalidade efetivar um direito fundamental**. Como se lê no Enunciado 115 do Conselho da Justiça Federal: “As limitações de direitos autorais estabelecidas nos arts. 46, 47 e 48 da Lei de direitos Autorais devem ser interpretadas extensivamente, em conformidade com os direitos fundamentais **e a função social da propriedade estabelecida no art. 5º, XXIII, da CF/88**” (Brasil, 2024a).

Isto sugere um movimento que, respeitadas as devidas diferenças, apresenta contornos familiares em relação a outros processos de criação e expansão dos direitos autorais, tais como no século XVIII, com a prensa, e nas décadas de 1990 e 2000, com a reprodução e compartilhamento online de arquivos. Assim como a Internet já representou “a morte do copyright” (Lunney, 2001), agora a IA representa “a morte da propriedade intelectual”.²⁵⁵ Nesse tipo de contexto, onde o potencial disruptivo da tecnologia ganha destaque, as limitações e exceções, enquanto instrumento de garantia do interesse público, acabam ganhando pouco espaço, especialmente diante do receio de que uma abertura permita processos de apropriação indevida (Ascensão, 2002; Valente, 2018).

O principal efeito dessa tensão é o reforço de uma visão polarizada e reducionista dos interesses envolvidos na legislação autoralista. Como visto no Capítulo 2, a proposta de “equilibrar” a proteção dos titulares com o interesse público termina por se render a uma lógica essencialista que falha em enxergar as necessidades da coletividade e desconsidera a realidade de países do Sul Global. Uma visão maximalista de direitos autorais exacerba esta questão ao reforçar a ideia de que a única forma de evitar uma “tragédia dos comuns” seria com a imposição de um direito de propriedade natural, absoluto, concentrado e privatista (Ostrom e Hess, 2007; Congost, Gelman e Santos, 2012). Consequentemente, qualquer pauta de Direito Autoral deverá, necessariamente, sujeitar-se a esta visão se não quiser ser vista com desconfiança, independentemente do quão relevante é o benefício pretendido para o interesse público – ou mesmo para autores e outros titulares, enquanto parte deste “público”.

Isto pôde ser visto à época da promulgação da Lei nº 9.610/98 no Brasil e é mais uma vez aparente no PL nº 2.338/2023, quando este estipula diversas restrições para o exercício da limitação ora proposta. Enxergam-se as limitações unicamente como brechas a serem exploradas por atores maliciosos, falhando em considerar a necessidade de viabilizar diversas atividades que não interferem com a exploração normal das obras ou com os interesses legítimos dos titulares – ou, se o fazem, é de forma mínima e *justificada* (como indica a regra dos três passos),

²⁵⁵ Para um exemplo, ver: “<https://www.newyorker.com/magazine/2024/01/22/who-owns-this-sentence-a-history-of-copyrights-and-wrongs-david-bellos-alexandre-montagu-book-review>”. Acesso em 09 out. 2024.

visando atender a um interesse público proeminente. Volta-se para a mesma noção de propriedade já criticada por Ostrom (1990; 2010), a qual opera em extremos: uma regulação concentrada nas mãos de poucos entes, mal ajustada à complexidade de práticas e interesses observados no dia-a-dia.

Notadamente, tal postura se mostra especialmente descolada das práticas e problemáticas apontadas por pesquisadores – os quais, lembre-se, gozam do mesmo status legal de autor do que qualquer artista e, portanto, fazem parte da mesma classe profissional que os direitos autorais se propõem a proteger. Entretanto, o ambiente de produção e divulgação científica, como visto no capítulo anterior, lida com questões ao mesmo tempo similares e distintas da indústria cultural: por um lado, para muitos cientistas, a garantia aos interesses sobre a exploração econômica ocupa papel secundário, sendo os interesses sobre a autoria e a integridade que parecem mais importar ao sucesso profissional e econômico (van Dalen e Henkins, 2012). Por outro lado, os levantamentos quantitativos levantados anteriormente (Tenopir et al, 2011, 2015, 2020; Andreoli-Versbach e Mueller-Langer, 2013; Wouters e Haak, 2015), bem como as entrevistas analisadas neste trabalho, mostram uma tensão constante entre buscar maior acesso a materiais para pesquisa e ter receio de compartilhar material e tê-lo apropriado por terceiros sem qualquer compensação ou crédito – o que reúne tanto os interesses de autores quanto o interesse público, algo impossível de conciliar dentro de um sistema que coloca estas questões como antagônicas, ao invés de complementares.

Ao mesmo tempo, o mercado editorial acadêmico também revela outras questões que o debate sobre a regulação de mineração de textos e dados no Brasil, hoje incorporado ao debate sobre regulamentação da IA, não parece enfrentar a fundo: conforme observado no Capítulo anterior, diversas editoras restringem o acesso a seu material, especialmente para fins de mineração de textos e dados, ao mesmo tempo em que demandam cessões de direitos sobre a produção dos pesquisadores e desenvolvem seus próprios sistemas de inteligência artificial – presumivelmente com o mesmo material que adquiriram de maneira lícita, sem nenhuma forma garantida de compensação adicional aos autores. Trata-se de um movimento duplo que limita e controla o compartilhamento e uso de dados por pesquisadores, enquanto amplia as possibilidades de abuso dos dados e materiais fornecidos por estes mesmos profissionais.

Assim, embora possa existir o receio de que uma limitação mais aberta ao compartilhamento e uso de materiais para fins de investigação ajudaria a legitimar uma apropriação desenfreada dos dados de pesquisa por estrangeiros, é pertinente considerar de que maneira o PL, em sua forma atual, se insere dentro das atuais relações de poder na academia –

especialmente em um ambiente no qual grandes editoras gozam de um crescente controle sobre dados graças a seu status e poder de mercado. Nesse sentido, observa-se a seção de direitos autorais e conexos do PL, apesar da maior restritividade quanto à mineração de textos e dados, traz elementos – como a falta de cobertura para pesquisadores não-vinculados a instituições, a ausência de critérios claros para a exigência de acesso lícito e a presença de previsões sobre liberdade negocial sem considerar assimetrias presentes no mercado – que reduzem sua capacidade de garantir segurança a cientistas, jornalistas, museus, arquivos e bibliotecas para que estes possam acessar e utilizar material para fins de interesse do país, enquanto pesquisadores no Norte Global gozam de maiores liberdades para aproveitar os dados fornecidos pelo Sul.

Desta forma, evidencia-se outro efeito deste processo de polarização no debate regulatório sobre mineração de textos e dados atualmente em curso no Brasil, bem como nas diversas regulações para IA em discussão ao redor do mundo: observa-se um cenário de “titulares contra desenvolvedores”, no qual as atenções visivelmente se voltam para as grandes empresas de tecnologia – as quais, como mostrado no Capítulo 1, se tornaram as principais envolvidas em litígios envolvendo o uso de obras protegidas no treinamento de inteligências artificiais. No entanto, a redução do debate a estes termos também corre o risco de ocultar as diversas relações de poder que permeiam a produção de cultura e conhecimento, especialmente entre autores e grandes titulares de direitos. Em diversos pontos, o atual debate regulatório sobre IA evidencia esse problema: não há considerações, no âmbito dos direitos autorais e conexos, de como se garantir que o cumprimento dos deveres de transparência e compensação seja revertido em benefício dos autores: ao contrário, as prerrogativas sobre o uso de IA são dadas aos titulares, a viabilidade do dever de transparência é deixada para um momento posterior e a regulação sobre remuneração é retirada das possibilidades de análise sob o *sandbox* regulatório, ao mesmo tempo em que a liberdade de negociação e o segredo comercial ocupam papel de destaque.

Tudo isto parece evocar um cenário muito parecido com o visto nos setores audiovisual e musical nas últimas décadas: uma caça desenfreada de serviços de compartilhamento de arquivo transformada em modelo de negócio, potencialmente retardando o surgimento das plataformas de *streaming* em vários anos (Carrier, 2012). Este processo eventualmente culminaria em modelos de negócio legítimos baseados em acordos entre titulares e plataformas digitais, mas a medida em que isto se reverteu efetivamente em favor dos artistas parece ser, na melhor das hipóteses, nebulosa: é o que indicam os debates sobre remuneração no ambiente digital na OMPI, que se intensificaram nos últimos anos (OMPI, 2015), os quais apontam para um crescente descompasso entre os lucros da indústria e os rendimentos dos artistas, potencializado por

questões de transparência sobre os valores repassados e o funcionamento de algoritmos de recomendação, além de uma complexa difusão de direitos ao longo da cadeia de valor (Butler, 2021; Calboli e Hwang, 2021; Castle e Feijóo, 2021; Sagna, 2021).

O campo da IA Generativa tem manifestado sinais desta dinâmica, com as tensões entre os profissionais criativos e o uso de inteligência artificial por grandes estúdios (e não por desenvolvedores) que constituíram um tema central nas greves de atores e roteiristas ocorridas nos EUA em 2023, conforme discutido no Capítulo 1. Trata-se de questões contratuais e laborais que perpassam o Direito Autoral, mas que não são facilmente resolvidas com uma simples atribuição de novas obrigações para desenvolvedores. Afinal, é necessário indagar em que condições ficam cientistas e artistas quando grandes titulares passam a fazer uso de inteligências artificiais, entram em acordos de licenciamento com empresas de IA ou desenvolvem, elas mesmas, seus próprios modelos – a exemplo das editoras acadêmicas vistas no Capítulo 4.

Ao mesmo tempo, observa-se outro problema: ao se adotar uma postura insular em relação ao restante do ordenamento jurídico, o Direito Autoral também é colocado como solução definitiva e suficiente para os desafios da IA Generativa, preocupando-se com uma expansão de direitos que não apenas ignora a função social da propriedade prevista no ordenamento jurídico pátrio e as implicações práticas para os profissionais afetados, mas também atua de forma desalinhada do ambiente de inovação e dos diversos desafios técnicos e regulatórios que influenciam a efetividade da regulação aprovada no Senado.

Nesse sentido, ressalta-se que o Brasil é um país que ainda apresenta pouca integração entre academia e setor privado, a despeito do potencial das instituições de pesquisa em gerar conhecimento capaz de ser convertido em bens e serviços de valor agregado. Como visto neste capítulo e no anterior, as políticas de inovação brasileiras, inclusive para IA, têm levado isto em consideração e tentado promover o desenvolvimento da universidade empreendedora. No entanto, a entrada de uma visão maximalista de Direito Autoral ao Projeto de Lei, reforçada institucionalmente, entra em rota de colisão com a estratégia de inovação assumida pelo país ao longo das últimas décadas. Isto se agrava quando se pensa na possibilidade de projetos de inteligência artificial desenvolvidos dentro de universidades: embora casos envolvendo IA Generativa mereçam uma ponderação e um cuidado à parte, o escopo assumido pelo Projeto de Lei em muito ultrapassa essas questões legítimas ao prever restrições e cobranças pelo desenvolvimento e uso de *qualsquer* inteligências artificiais – mesmo as que, por natureza, não têm o mesmo efeito substitutivo aos autores que a GenAI possui.

Por outro lado, também se constata que, ao mesmo tempo em que se aplica um escopo deveras amplo para a proteção de obras, também restringe-se o tipo de mineração de textos e dados que poderia ser conduzida por instituições de pesquisa, jornalismo e conservação. Parte da razão disso está no fato de que a discussão regulatória, hoje, está voltada especificamente para a *inteligência artificial* – ainda que a aplicação da mineração de textos e dados, como observado no Capítulo 1, vá para além disto. O dispositivo previsto no parecer final da CJSU-BIA, que permitia a mineração para outras operações com IA, poderia ser utilizado como uma forma, dentro da lei, de ao menos permitir o uso de certas ferramentas automatizadas, a exemplo de *bots* para coleta de dados não-pessoais e referencial teórico. No entanto, a ausência desse tipo de disposição retira até mesmo esta possibilidade, trazendo garantias extremamente limitadas a um número reduzido de profissionais. Isto evidencia outro problema hoje presente no debate sobre inteligência artificial no Direito Autoral: é por este tema específico que hoje se pauta o tema da mineração de textos e dados no ambiente regulatório brasileiro, uma vez que o temor em relação aos modelos de GenAI parece ser o único ponto de interesse do regulador. Exclui-se, abertamente, qualquer outra forma de coleta e análise de dados, que segue em solo juridicamente instável.

Outro efeito deste processo de reducionismo está na ausência de uma reflexão mais profunda acerca dos desafios técnicos envolvidos nos processos de transparência, identificação e fiscalização das obrigações legais atualmente sob análise do Legislativo. Em que pese a constante evolução das ferramentas de identificação de obras, não apenas a mineração de textos e dados para o treinamento de IA tende a lidar com volumes massivos de material, como a própria verificação da presença e usos diretos de obras específicas em modelos (generativos ou não) ainda não parece ser aferível com total segurança (Guadamuz, 2023, 2024). Diante deste cenário e à luz das restrições impostas por medidas anti-burla (Dusollier, 2005; Derclaye, 2008), questiona-se como os desenvolvedores serão capazes de cumprir as exigências legais ora prescritas e como os titulares conseguirão verificar possíveis violações – algo que, pelo que indicam muitos dos processos judiciais descritos no Capítulo 1, segue sendo uma tarefa de duvidosa exequibilidade.

Nessa mesma toada, também cumpre observar as implicações deste tipo de regulação para o uso de dados *de* pesquisa e *para* pesquisa. Como vimos no Capítulo 1, é bem verdade que a jurisprudência brasileira remove esse tipo de material da seara do Direito Autoral, mas eles estão, muitas vezes, associados a trabalhos científicos que *são* protegidos. Como separar estes elementos um do outro? Como definir a titularidade desse material nessa situação? No

caso de pesquisas conduzidas pelo Estado ou por ele apoiadas, como garantir juridicamente que este material seja disponibilizado ao público, tal como defendido pelas pessoas entrevistadas neste trabalho? Todas estas questões se colocam como importantes para resolver, ao menos em parte, a atual incerteza regulatória que assombra pesquisadores e instituições de produção e preservação do conhecimento.

Até o momento, entretanto, o que se vislumbra atualmente é um ambiente polarizado, dividido entre proponentes de uma intervenção mínima em nome da inovação, de um lado, e, de outro, uma postura estatal marcada por uma visão maximalista de Direito Autoral, motivada por temores legítimos acerca dos impactos da IA Generativa para a atividade criativa. Em ambos os casos, contudo, a presença de um Direito Autoral calcado em um modelo de apropriação privatista e absoluto segue sem questionamento, de forma que considerações quanto ao interesse público parecem ocupar papel secundário. O resultado são dispositivos legais descolados das necessidades dos pesquisadores, das características das tecnologias envolvidas, do papel da pesquisa no ambiente de inovação e do próprio ordenamento jurídico brasileiro – e com eficácia limitada para restringir um processo de concentração de poder nas mãos de grandes titulares e plataformas, em prejuízo tanto dos autores como da sociedade em um âmbito geral.

5.6. Repensando o paradigma regulatório

Diante do cenário que se coloca para a regulação da mineração de textos e dados no Brasil, entende-se que o momento é propício para uma reanálise das disposições sobre Direito Autoral aqui discutidas e como estas podem melhor se ajustar às necessidades do ecossistema de pesquisa – seja científica, jornalística ou para fins de conservação. De forma mais contundente, observa-se a necessidade de se pensar em modelos regulatórios que se afastem de um paradigma estritamente privatista de propriedade e situem os direitos autorais não como isolados ou opostos ao interesse público, mas como parte deste. Trata-se de um sistema que compreenda os diferentes interesses envolvidos (como autores, pesquisadores, jornalistas, profissionais do ramo da conservação, titulares, instituições de pesquisa, desenvolvedores e sociedade civil) enquanto partes integradas e complementares, guiadas por uma visão regulatória voltada para a proteção do trabalho artístico e científico e pelo acesso ao conhecimento.

Para tanto, também é importante considerar opções regulatórias que compreendam o Direito Autoral não como um sistema fechado e de limites rígidos, mas como um sistema que oferece diversas possibilidades regulatórias que podem ser mobilizadas em prol das necessidades locais. Nesse sentido, destaca-se que mesmo a regra dos três passos, por seu caráter principiológico, serve meramente como guia para a elaboração de usos livres – estabelecendo não

apenas limites, sim, mas também um espaço amplo para regulações ajustadas à realidade de cada membro da Convenção de Berna. Com efeito, o Capítulo 3 expôs diversas. Como exemplo, convém resgatar o exemplo da legislação japonesa acerca da mineração de textos e dados, que delimita hipóteses legais onde o uso de obras protegidas é permitido, porém sem se prender a uma série de imposições com receio da má apropriação de conteúdos – com a exigência de aplicação não-expressiva servindo como um norte para o estabelecimento do “certo caso especial” (Ueno, 2021).

Por outro lado, também é crucial que a delimitação dos direitos dos titulares e das limitações e exceções também siga critérios claros e factíveis. Por exemplo, é importante que se estabeleça o que contaria como “acesso lícito” e “exploração normal da obra”, ou qual o nível de especificidade necessário para o cumprimento da cláusula de transparência (é preciso indicar obra por obra? Qual o detalhamento a ser dado em relação ao processo de coleta?). Estas são questões que podem não ter solução simples – e que a experiência de um *sandbox* regulatório ou outra instância de atuação estatal poderia ajudar a elucidar.

Ademais, também é importante considerar a variedade de modelos regulatórios já existentes ao redor do mundo sobre este tema, ponderando-se as forças e fraquezas de cada modelo e como estas podem ser aplicadas à realidade institucional e material brasileira. Nesse sentido, observa-se que, embora propostas regulatórias anteriores tenham empreendido um esforço em combinar elementos de leis diferentes (como a Diretiva CDSM e a Lei de Direitos Autorais do Japão), foi-se dando lugar a um modelo que em grande parte emula o que se vê na União Europeia, sem que se atente para as diferenças entre Europa e Brasil: ignora-se, por exemplo, que iniciativas como o *AI Act* existem, por exemplo, em um contexto no qual limitações e exceções para mineração de textos e dados na pesquisa já foram objeto de discussão e regulação anterior. Não apenas isto nunca ocorreu no Brasil, como a Lei de Direito Autoral brasileira possui direitos exclusivos que interferem diretamente com esses processos tecnológicos.

Assim, alcançar uma regulação do Direito Autoral que vá ao encontro do interesse público também implica em ser capaz de olhar para os lados, buscar aprendizados com as mais diversas experiências regulatórias e analisar o que tem serventia, o que é inadequado e o que pode ser adaptado para a realidade local. Dentre os modelos apresentados no Capítulo 3, pode-se destacar diversos exemplos que podem servir como inspiração: as definições legais japonesas sobre uso não-expressivo e não-interferência na exploração normal da obra; as disposições, na Diretiva CDSM, sobre armazenamento seguro e manutenção de cópias para verificabilidade dos dados de pesquisa; entre outros.

Simultaneamente, é importante voltar-se para dentro e compreender os limites e possibilidades advindos do próprio Direito brasileiro. Nesse sentido, o Direito Autoral, enquanto associado ao direito de propriedade, deve observância não somente aos outros direitos fundamentais, mas à função social da propriedade prevista na Constituição Federal (Souza, A. 2005; Lewicki, 2007). Por outro lado, deve-se reconhecer as restrições legais já existentes na lei em relação à coleta e análise de dados com obras protegidas, como o art. 29, X da LDA, bem como o efeito dos instrumentos legais anti-burla no art. 107 da mesma Lei sobre as limitações e exceções. Ao mesmo tempo, convém pensar disposições inexistentes em outros locais, mas capazes de agregar maior serventia ao interesse público – como, por exemplo, a extensão da limitação para instituições de pesquisa a entidades jornalísticas, que foi retirada do texto final do PL no Senado quando sua manutenção poderia servir a ferramentas como serviços de checagem de fatos, que podem utilizar mecanismos automatizados como auxílio ao trabalho jornalístico (Aos Fatos, 2021; 2024).

Portanto, uma legislação adequada para a mineração de textos e dados diante do ordenamento jurídico pátrio se beneficiaria, essencialmente, dos seguintes pontos: entender as possibilidades e limites regulatórios do Direito Autoral; estabelecer mais claramente as exigências jurídicas a serem seguidas pelos agentes envolvidos no processo de mineração de textos e dados (titulares, pesquisadores, desenvolvedores e outros); observar os diversos modelos regulatórios existentes ao redor do mundo, observar seu funcionamento, os dispositivos que os diferenciam e usá-lo como exemplo e, por fim, conciliar este aprendizado com os princípios e normas internas, aproveitando possibilidades ainda não exploradas de beneficiamento do interesse público.

De forma mais crucial, porém, uma abordagem sensível e adequada às atividades de pesquisa requer que o debate regulatório em curso considere pragmaticamente como o Direito Autoral influencia o trabalho de pesquisa. Mais especificamente, urge a promoção de um maior diálogo com estes agentes no debate regulatório – especialmente profissionais e instituições que atuam não apenas no ambiente científico, como pesquisadores, instituições de pesquisa, repositórios e editores científicos, mas também em outras atividades de investigação e documentação, como jornalistas, museus, arquivos e bibliotecas.

Embora não haja necessariamente uma resposta definitiva para esta questão, entende-se vital que ela envolva alguns pontos em particular. Primeiramente, é necessário repensar a restrição da limitação para desenvolvimento e uso de ferramentas automatizadas a instituições de pesquisa, museus e arquivos e considerar as condições de trabalho de profissionais não-vinculados a instituições de pesquisa, os quais necessitam ter algum tipo de resguardo que

permita a manutenção de sua produção científica e facilite, eventualmente, sua recolocação profissional. Ao mesmo tempo, a restrição da limitação para pesquisa a entes institucionais já impõe a proibição de atividades com intuito de lucro ou a atuação de instituições controladas por atores comerciais – de forma que a exclusão de pessoas físicas, mantidas estas restrições comerciais, não parece servir a nenhuma finalidade além de concentrar a capacidade e autoridade na produção científica nas mãos de poucos. Desta forma, uma limitação não para instituições de pesquisa, mas para fins de pesquisa parece se apresentar como uma opção regulatória mais coerente.

Outro ponto não menos importante para o debate diz respeito a qual tipo de pesquisa deve ser considerado na regulação. Afinal, cumpre lembrar, conforme visto no Capítulo 1, que a mineração de textos e dados não está restrita ao desenvolvimento de inteligências artificiais: na verdade, precede estes sistemas e conta com diversas aplicações que não têm como objetivo o desenvolvimento de IA, mas compõem um universo muito maior de atividades que têm em comum simplesmente a busca por métodos de análise de dados mais aprimorados. Assim, restringir-se à questão mais imediata da IA Generativa é uma decisão que implica em manter diversos pesquisadores em um cenário praticamente inalterado de incerteza regulatória. Note-se que a versão da limitação para pesquisa em IA proposta pela CJSUBIA continha, no antigo art. 42, §2º, uma permissão para atividades analíticas usando sistemas de IA (Brasil, 2022c),²⁵⁶ o que poderia ser interpretado como um avanço nesse sentido.

Conectar-se com a realidade material das atividades de pesquisa também requer conciliar as demandas dos titulares com o incentivo ao ambiente de inovação e desenvolvimento – o que, no Brasil, envolve o incentivo às relações universidade-empresa. Restringir limitações pra mineração de textos e dados a fins não-comerciais pode encontrar eco em argumentações de compensação financeira, mas a efetivação deste dispositivo necessita considerar, por exemplo, como isto afeta processos de licenciamento gerenciados por núcleos de inovação tecnológica ou a atuação das incubadoras caso um grupo decida transformar seu projeto em um produto ou serviço. Resolver estas questões exige a superação do atual embate entre Direito Autoral restritivo e visões com mais foco na inovação presente nas emendas ao PL nº 2.338/2023 analisadas neste trabalho. Aqui, entende-se que o segredo reside no diálogo, conciliando diferentes responsabilidades e projetos do Estado e entender a propriedade intelectual não como um

²⁵⁶ “Art. 42 (...) § 2º Aplica-se o disposto no caput à atividade de mineração de dados e textos para outras atividades analíticas em sistemas de inteligência artificial, cumpridas as condições dos incisos do caput e do § 1o, desde que as atividades não comuniquem a obra ao público e que o acesso às obras tenha se dado de forma legítima.” (Brasil, 2022c).

elemento separado do ambiente político, mas sim como parte de um projeto maior de inovação e desenvolvimento. Nesse sentido, percebe-se que muitas das regulações referentes à mineração de textos e dados observadas no Capítulo 3 se encontram situadas dentro de uma visão mais ampla de promoção do desenvolvimento tecnológico, e não à parte desta – a exemplo das considerações sobre fomento a novos mercados no caso Sony Betamax (Estados Unidos, 1984), da iniciativa Society 5.0 (Japão, 2017a) e do Mercado Único Digital (União Europeia, 2019b).

Finalmente, o rompimento com a polarização hoje presente no debate sobre IA e com o paradigma tradicionalista hoje visto na proposta legal em discussão também implica em reconhecer as possibilidades de atuação dos entes interessados para além da esfera legal balizada pelo Projeto de Lei – seja no sentido de garantir o seu cumprimento, seja para estabelecer modelos que se oponham à concentração de poder dentro da academia. Um exemplo disto está no cumprimento da imposição de regras de transparência do art. 62, o que requererá, dentro da comunidade acadêmica, a implementação de padrões internos de documentação do material de pesquisa. Além disto, o fato de a limitação do art. 63 não se estender a medidas tecnológicas de proteção já pode forçar pesquisadores e instituições de pesquisa a buscar o licenciamento do material de pesquisa com os titulares – o que, na ausência do *sandbox* regulatório, significará uma regulação pela via privada. Diante do aparente estado de dissenso em que se encontra a academia no que diz respeito ao tratamento e compartilhamento dos dados de pesquisa, isto aponta para a necessidade de uma maior organização das instituições de pesquisa – e entre instituições, financiadores e pesquisadores – no sentido de instituir e harmonizar diretrizes e protocolos gerais de gestão de dados.

Ao mesmo tempo, um caminho alternativo nesse sentido é apontado pela Wikimedia Enterprise, projeto pertencente à Wikimedia Foundation (2025) voltado para fins comerciais abertos: através de um recente acordo com a *startup* Pleias, ambas as entidades declararam ter firmado uma parceria no sentido de utilizar o acervo da Wikimedia como parte de um projeto para o desenvolvimento de bases de dados devidamente estruturadas, legalizadas e abertas como forma de atender às exigências do *AI Act* europeu (Wikimedia Enterprise, 2025). Tal iniciativa aponta para o uso de regras de transparência como incentivo ao desenvolvimento de modelos e repositórios abertos – o que, embora não anule as diversas questões anteriormente levantadas sobre a efetivação das exigências legais, representa uma opção que se afasta do maximalismo em benefício da abertura de dados e do acesso ao conhecimento.

Por outro lado, também é crucial pensar em mecanismos capazes de reduzir a possibilidade de que sistemas de gestão mais abertos terminem servindo, inadvertidamente, a

processos de concentração de conhecimento e capital. Mais especificamente, Widder, Whittaker e West (2024) observam que, em um cenário no qual grandes processos de análise de dados requerem uma enormidade de recursos e mão-de-obra e a infraestrutura material e informacional está concentrada nas mãos de poucos atores privados, cria-se um ambiente propício para a cooptação de projetos abertos por grandes empresas de tecnologia e, em última análise, a perpetuação de uma relação colonial de captura de recursos do Sul Global por atores do Norte. Assim, embora pareçam uma alternativa à primeira vista mais benéfica do que um modelo fechado de propriedade, os autores sugerem que modelos abertos devem também ter suas possibilidades e desafios abordadas com o devido escrutínio e com regras de governança apropriadas, ao invés de serem acriticamente adotados.

Neste sentido, as entrevistas e documentos analisados ao longo deste trabalho, especialmente no Capítulo 4, sugerem que parte dos dilemas encontrados estão relacionados ao poder de mercado de grandes editoras e plataformas em relação ao controle sobre os dados de pesquisa, potencializado pela demanda por produções atreladas a publicações de alto valor de impacto. Nesse cenário, a ausência de disposições sobre como efetivamente compensar os autores pelo uso de obras em sistemas de IA representa um desafio para a redução dessas desigualdades, uma vez que deposita diversos encargos ao pesquisador sem impor contrapartidas aos grandes titulares.

Este exemplo também ilustra a pertinência de mecanismos que propiciem alternativas ao modelo de relações entre pesquisadores, instituições de pesquisa e periódicos, com práticas e modelos que se oponham à apropriação do conhecimento e à deturpação da ciência aberta. Exemplos incluem regimes de licenciamento como Creative Commons (mencionado tanto nos estudos quantitativos como nas entrevistas deste trabalho) e os termos de uso da plataforma GISAID (mencionada no Capítulo 4), bem como a postura manifestada recentemente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (2024), no sentido de modificar os padrões de avaliação para minimizar o peso do fator de impacto. Aqui, a ideia central é facilitar o acesso a materiais de pesquisa e, ao mesmo tempo, resguardar autores contra a má apropriação de suas pesquisas e dados subjacentes, a fim de lhes garantir maior confiança para compartilhar seu trabalho.

Ao fim e ao cabo, entende-se que, por trás destas considerações, existe uma questão essencialmente teórica em relação ao modo de se entender a propriedade e sua função. Mais especificamente, defende-se que uma proposta regulatória adequada para a mineração de textos e dados na pesquisa, passa, necessariamente, pelo rompimento com o paradigma de “proteção

versus abertura” ou, nos termos de Ostrom (1990), “público versus privado”, onde a centralização de prerrogativas (seja no Estado, seja em agentes privados específicos) se coloca como a única forma de gestão possível, em uma tentativa de evitar uma hipotética “tragédia dos comuns” sobre um bem que cresce ao invés de se exaurir. No lugar disto, propõe-se a adoção de um modelo de “feixe de direitos”, que compreenda a possibilidade de formas mais descentralizadas e híbridas de gestão da propriedade, que reconheçam os direitos de autor como ferramenta importante de estímulo ao crescimento do conhecimento, desde que comprometida com a acessibilidade e difusão desse conhecimento, Isto não implica no uso de modelos irrestritamente abertos, mas sim de regimes coletivos, onde os diferentes atores envolvidos na produção e circulação do conhecimento são considerados de forma mais ativa na gestão dos *knowledge commons*.

Propõe-se, assim, uma inversão da lógica de “proteger o máximo possível”, onde o Direito Autoral é a regra e o atendimento à coletividade, a exceção, para um modelo centrado em “abrir da melhor forma possível”, onde as necessidades do interesse público são a regra geral que norteia o Direito Autoral, como caminho mais adequado para fomentar e cultivar a produção intelectual e evitar a apropriação indevida dos *knowledge commons* (Argyres e Liebeskind, 1998; Nelson, 2003; Orsi e Coriat, 2006; Ostrom e Hess, 2007b; Hess, 2008; Coriat, 2013) – seja por parte de grandes titulares, seja por parte de empresas de tecnologia.

Em suma, observa-se que o Brasil já não passa incólume ao debate público sobre mineração de textos e dados e sua interseção com o Direito Autoral. No entanto, os recentes desdobramentos no campo da inteligência artificial parecem ter trazido à tona um debate regulatório polarizado e um posicionamento estatal conservador, levando a um modelo regulatório descolado das necessidades reais dos pesquisadores, do ordenamento jurídico brasileiro e do projeto de desenvolvimento assumido pelo próprio Estado – terminando por privilegiar os interesses de grandes titulares ao invés de facilitar uma mediação equilibrada entre o incentivo à produção de conhecimento e a garantia de que a sociedade, enquanto destinatária deste bem, possa acessá-lo e usá-lo como base para produzir mais conhecimento. A menos que este paradigma seja debatido e desconstruído, portanto, corremos o risco de não somente caminhar em círculos, mas também em uma espiral rumo a uma maior comoditização do conhecimento, a um abismo mais profundo entre Norte e Sul Global na produção de conhecimento e inovação, à relegação dos países em desenvolvimento à posição de fornecedores de insumos de pesquisa e à repetição de promessas vazias aos autores – sejam eles artistas ou cientistas.

CONSIDERAÇÕES FINAIS

Este trabalho buscou avaliar se o debate e proposta regulatórios sobre a mineração de textos e dados no Direito Autoral atendem às necessidades do ecossistema de pesquisa, tendo-se em vista o ordenamento jurídico pátrio e as políticas públicas de inovação presentes no contexto brasileiro. Mais especificamente, procurou-se investigar os usos e aplicações da mineração de textos e dados no ambiente de pesquisa e analisar a estrutura e a função do sistema de proteção a bancos de dados e o acesso ao seu conteúdo no Brasil, sobretudo pelo ponto de vista do Direito Autoral.

Neste processo de análise, observou-se, no Capítulo 1, que o uso das tecnologias de mineração de textos e dados tem ocupado posição de destaque no processo de inovação e já constitui prática corriqueira e essencial na pesquisa, com aplicações que vão muito além da atividade científica – alcançando, também, a investigação jornalística e atividades de conservação. No entanto, constata-se a presença de um cenário de incerteza regulatória no que diz respeito à possibilidade de emprego dessas ferramentas quando estas envolvem o uso de obras protegidas por Direito Autoral – especialmente considerando a consolidação, nas últimas décadas, de uma estrutura jurídica e tecnológica que permite o controle sobre o acesso a bancos de dados e ao material neles contido, ainda que este não seja protegido por direitos autorais. Também se observa que a crescente adoção social e econômica de modelos de inteligência artificial, que utilizam técnicas de mineração para o seu desenvolvimento e operação, tem provocado pressões maiores pelo reforço deste sistema de proteção – motivadas, particularmente, pelos impactos gerados pela IA Generativa aos titulares de direitos.

Neste contexto, ressalta-se a importância de instrumentos regulatórios que garantam o alinhamento do Direito Autoral à sua função social de fomento à expansão do conhecimento e da cultura, conciliando o incentivo à produção intelectual com o interesse público relacionado às possibilidades de compartilhamento deste material para a produção de novos conhecimentos. Nesse sentido, como visto no Capítulo 2, as limitações e exceções aos direitos autorais cumprem um papel vital enquanto expressão desse alinhamento, servindo de base para a discussão de medidas de fomento à pesquisa no âmbito supranacional. Ao mesmo tempo, contudo, constatou-se a manutenção de um paradigma teórico de Direito Autoral que privilegia a propriedade privada como modelo ideal de gestão, colocando os interesses privados de maneira mais definida e consolidada em detrimento dos interesses da coletividade – limitando a efetividade desses mecanismos de adequação e colocando a inclusão da mineração de textos e dados para

pesquisa em posição delicada diante do aumento de pressões para a restrição de técnicas de coleta e análise.

Nesse sentido, diversas jurisdições têm buscado se dirigir à questão da mineração de textos e dados de diferentes maneiras, apresentando uma diversidade de opções regulatórias que oferece múltiplas oportunidades de aprendizado: desde o modelo casuístico de cláusula aberta exemplificado pelo *fair use* estadunidense, passando por usos livres destinados a fins de pesquisa de forma geral, até as limitações japonesas para usos não-expressivos, os usos livres específicos para fins acadêmicos e para propósitos mais gerais da Diretiva CDSM da União Europeia e, por fim, as limitações para MTD no *fair dealing* inglês. Mesmo na América Latina, habitualmente mais restritiva em relação as limitações e exceções, alguns países já começaram a debater este tópico em caráter oficial. Em todos esses exemplos, nota-se uma tensão entre a proteção dos interesses de titulares e um crescente interesse estatal de fomentar a pesquisa intensiva em dados dentro de uma estratégia de incentivo à inovação tecnológica.

Esta ligação entre pesquisa e inovação não se trata de um detalhe trivial, uma vez que, como abordado no Capítulo 4, as práticas de gestão de dados e propriedade intelectual no ecossistema de pesquisa vêm constituindo elemento de destaque nas políticas de inovação. No entanto, este ambiente apresenta diversos paradoxos que impõem desafios à regulação da mineração de textos e dados: observam-se políticas governamentais de estímulo ao empreendedorismo universitário, à parceria entre empresas e ao reforço da propriedade intelectual, ao mesmo tempo em que se vê o crescimento de movimentos pelo acesso aberto, inclusive de dados. De forma paralela, observa-se, por parte de pesquisadores, jornalistas e profissionais de conservação, uma preocupação em relação às possibilidades de acesso e uso de materiais para a realização de suas atividades, ao mesmo tempo em que manifestam temores de que uma abertura indiscriminada facilite um processo de apropriação do conhecimento por grandes titulares e por países do Norte Global em detrimento do Sul. Tudo isto é potencializado pela enorme variedade nos tipos de dados usados na pesquisa e por uma enorme incerteza regulatória causada por um labirinto de termos de uso e pela falta de padrões unificados sobre gestão de dados.

Diante deste cenário, o Capítulo 5 constata que, em que pese a inclusão da mineração de textos e dados nas mais recentes estratégias de inovação no Brasil, são as atividades legislativas envolvendo a regulação da inteligência artificial as principais responsáveis por promover maiores discussões sobre mineração de textos e dados no âmbito estatal. Neste contexto, nota-se uma crescente tensão entre esforços para desregulamentação das tecnologias de IA – facilitando a apropriação e uso irrestrito de dados por empresas de tecnologia em detrimento do bem-

estar coletivo – e, por outro lado, o avanço de uma visão maximalista de Direito Autoral, que desconsidera os usos mais amplos da mineração de textos e dados, a função social da propriedade e a realidade material dos agentes envolvidos na produção de conhecimento. Por trás destas posições, observa-se a manutenção de um modelo de propriedade que privilegia os interesses privados em detrimento do interesse público – aprofundando a precarização do trabalho de pesquisadores, jornalistas e entidades de conservação e, ao mesmo tempo, acentuando a concentração de poder econômico e informacional por grandes titulares de direitos e empresas de tecnologia de porte à revelia de autores, usuários e da sociedade em geral.

Desta maneira, o principal resultado encontrado neste trabalho é a constatação de que, embora o debate público sobre a proteção de bases de dados, bem como a mineração de textos e dados no Direito Autoral, tenha se expandido de forma considerável nos últimos anos, esta discussão vem sendo profundamente impactada por um processo cada vez mais visível de subsunção da gestão do conhecimento a uma visão privatista de propriedade, pouco compatível com as demandas impostas pelo interesse público.

Mais especificamente, observa-se que as primeiras regulações sobre mineração de textos e dados no Direito Autoral datam dos anos 2000 e 2010 – épocas que testemunharam uma maior simpatia pelo interesse público e pelo uso das L&E como instrumento de promoção do desenvolvimento intelectual e econômico. Foi nesta época que se firmaram iniciativas como o Tratado de Marrakesh, em sede internacional, o Mercado Único Europeu, na União Europeia, e, no Brasil, a criação de um arcabouço jurídico mais expansivo para os usos livres. Entretanto, o debate sobre a análise de dados para fins de pesquisa no Direito Autoral pátrio se encontra em meio a um cenário que, em certos aspectos, lembra o “fim do copyright” do final da década de 1990: o medo do potencial disruptivo da tecnologia e a expansão dos direitos de propriedade intelectual como forma de proteger os interesses dos titulares em meio a este cenário – à revelia não apenas dos usuários, mas também dos próprios artistas.

Este processo de recrudescimento, nota-se, termina por resgatar uma visão de Direito Autoral enquanto direito natural, absoluto e monolítico, de forma que o atendimento aos interesses dos titulares de direitos se coloca como um fim em si mesmo ao invés de um meio para o atendimento ao interesse público, de forma que a observância a outros direitos fundamentais passa, na prática, a ser vista com desconfiança. Também se nota uma clara tentativa de aproximação regulatória do modelo mais restritivo de regulação observado na União Europeia, sem uma devida ponderação sobre a adequação dessas regras ao ordenamento jurídico brasileiro – especialmente no tocante à função social da propriedade, imposta pela Constituição

O resultado deste processo até o momento é um debate regulatório amplamente tomado por uma visão polarizada do Direito Autoral, reproduzindo uma noção abstrata de propriedade descolada do contexto histórico, do ordenamento jurídico e da realidade material dos agentes envolvidos na produção de conhecimento e cultura. Conseqüentemente, este processo de expansão dos direitos dos titulares se mostra paradoxalmente reducionista em múltiplos sentidos: primeiramente, coloca-se o instituto como uma disputa entre “titulares versus empresas de tecnologia”, impondo inúmeras restrições à atividade acadêmica ao mesmo tempo em que estabelecem amplas prerrogativas com retorno incerto aos autores. Em segundo lugar, adere-se a uma interpretação restritiva da regra dos três passos, ignorando seu propósito não como um comando, mas como um guia. De forma similar, a emulação evidente da regulação da União Europeia ignora não apenas a multiplicidade de experiências regulatórias de mineração de textos e dados ao redor do mundo, mas também os próprios elementos internos ao Direito Autoral brasileiro que nos diferenciam das demais regiões do planeta – replicando estruturas institucionais que meramente perpetuam assimetrias de poder entre autores e sociedade no Sul Global e os grandes titulares do Norte.

De forma mais contundente, porém, esta visão reducionista se mostra incapaz de dar conta da complexidade de experiências vivenciadas pelos pesquisadores, os quais manifestam, simultaneamente, interesses enquanto titulares e membros do público. Mais especificamente, ignora-se a existência de um ambiente cujos profissionais ainda se queixam de questões básicas de acesso para a efetivação dos seus deveres institucionais, ao mesmo tempo em que permanecem desprotegidos contra a apropriação de sua produção por grandes titulares estrangeiros. Nesse sentido, a falta de maior clareza regulatória sobre usos de obras protegidas para fins de pesquisa, além da insistência em impor um mesmo regime geral de direitos e obrigações para todas as obras e mercados, desconsidera a grande diversidade de comunidades, práticas de coleta, gestão e uso de dados, paradigmas e mesmo de materiais existentes no ambiente de pesquisa. Impede-se, portanto, a construção de uma regulação que efetivamente considere a participação dos responsáveis pelas atividades de pesquisa – sejam eles pesquisadores, jornalistas ou profissionais no ramo da conservação.

Definir soluções em termos regulatórios para todos estes problemas não é tarefa simples: afinal, como dito anteriormente, o ordenamento jurídico brasileiro não opera em um vácuo, mas é inserido em um contexto internacional marcado pela predominância de uma visão liberal de Direito Autoral, pelo poder de mercado de grandes *players*, tanto na indústria cultural como no setor de tecnologia, e por crescentes preocupações com o efeito substitutivo da

inteligência artificial sobre profissionais da arte. Por outro lado, as questões de diversidade acima e as críticas da literatura sobre gestão dos *commons* alertam para a improbabilidade de uma única solução possível.

Entretanto, este trabalho conclui que o principal problema a ser solucionado neste cenário não envolve necessariamente a adoção de um texto específico – porque é, antes de tudo, um problema de ordem teórica. Considera-se, portanto, que o alcance de qualquer solução regulatória minimamente adequada às necessidades dos pesquisadores no âmbito da mineração de textos e dados requer um rompimento com o paradigma tradicionalista de Direito Autoral, reconhecendo a historicidade do instituto, os múltiplos atores e mercados envolvidos e, acima de tudo, a interação do instituto com outros ramos do Direito e seu papel dentro de um cenário mais amplo de fomento ao interesse público. Tal escolha, entende-se, não vem para minimizar a importância dos direitos de autor e conexos: pelo contrário, visa uma distribuição mais eficaz de prerrogativas, a fim de trazer maior resguardo aos autores e uma maior participação da sociedade na produção do conhecimento.

No tocante à regulação da mineração de textos e dados e inteligência artificial no Brasil, isto implica na necessidade de se discutir a capacidade de o texto proposto no momento trazer mais segurança e possibilidades para pesquisadores brasileiros, na sua condição de profissionais que produzem conhecimento e querem protegê-lo, ao mesmo tempo em que precisam acessá-lo com mais facilidade. Alguns pontos pertinentes para isto incluem:

- A possibilidade de uma regulação centrada em *instituições*, em favor de um uso livre para *fins* de pesquisa científica e jornalística e preservação, ao menos sem fins lucrativos. O propósito é duplo: incluir pesquisadores sem vínculo institucional, que precisam manter sua produção para ingresso em uma instituição, e compensar o problema de sobreinclusão no uso do termo “inteligência artificial” para cobrir processos automatizados de mineração de textos e dados que não envolvam o treinamento de IA;
- Nesse mesmo sentido, a permissão para o uso em outros processos de análise que não o treinamento dos sistemas de IA ou, alternativamente, uma cláusula específica para o uso de ferramentas automatizadas de coleta e análise de dados – obedecidas as demais exigências do artigo – ajudaria a evitar o esvaziamento da limitação;
- Uma discussão mais aprofundada quanto à manutenção da exigência de acesso lícito e sua definição legal (por exemplo: “considera-se o acesso lícito quando o conteúdo estiver disponibilizado ao público”), permitiria um controle e verificação

mais factíveis sobre o material de pesquisa, trazendo mais benefícios aos pesquisadores sem trazer prejuízos aos titulares, uma vez que a limitação dispõe de outras ferramentas muito mais eficazes em excluir a IA Generativa ou outros usos danosos aos artistas;

- Uma discussão mais clara em relação aos interesses não-patrimoniais e ao incentivo a modelos de compartilhamento que valorizem o crédito à fonte dos materiais coletados;
- Definições mais específicas sobre o formato exigido ou aceito para que a reserva legal ao uso de obras em IA seja considerada válida: seja por meio de mecanismos legíveis por máquina (como comandos automáticos de negação de raspagem ou outras ferramentas de gestão de direitos/MTP), por meio de contratos ou mediante declarações públicas e, além disso, se a proibição retroagirá para usos feitos antes da proibição – o que, como visto, pode ser complexo de efetivar;
- Ajustes em relação às referências quanto aos beneficiários das obrigações de transparência e remuneração. Mais especificamente, a reiteração do termo “autor” ao invés de “titular” reduziria a dubiedade na redação das propostas hoje em curso, colocando-as em acordo com os demais direitos exclusivos da LDA ao reafirmar o autor como titular originário e impedir a apropriação de prerrogativas por titulares sem consentimento dos autores;
- Neste mesmo sentido, um debate maior sobre regras de remuneração justa aos autores, a fim de garantir que os criadores dos materiais utilizados possam efetivamente participar e se beneficiar do processo de licenciamento de obras e evitar acordos fechados entre grandes titulares e empresas de tecnologia. Ao mesmo tempo, é essencial que se discuta quais os possíveis modelos a serem empregados para esta remuneração;
- Ponderações mais detalhadas sobre como conciliar o Direito Autoral com o incentivo à universidade empreendedora e às relações universidade-empresa, que compõem eixos importantes da política de inovação brasileira;
- De forma mais geral, uma abordagem menos restritiva em relação à regra dos três passos, qualificando, de forma mais precisa, a “exploração normal da obra” dentro dos limites de uso da expressão contida nas obras; bem como reiterando o papel dos “interesses legítimos dos titulares” ao invés de meramente os “interesses econômicos”;

- Por fim, a exploração de uma variedade maior de possibilidades regulatórias, tomando os diferentes modelos adotados para regulamentação da mineração de textos e dados e de IA ao redor do mundo enquanto oportunidades de aprendizado, analisando elementos de destaque, forças e fraquezas e como melhor aproveitá-las no contexto nacional.

Também é forçoso reconhecer que o alcance de um sistema efetivamente benéfico ao interesse público envolvido na mineração de dados e textos para pesquisa não se restringe ao escopo da discussão legislativa em torno de projetos de lei específicos para inteligência artificial. Isto porque nenhuma proposta de regulação de Direito Autoral verdadeiramente comprometida com a Constituição e com a finalidade social do instituto de promoção do conhecimento e cultura pode escapar da necessidade de pensar as limitações e exceções como mais do que simples usos livres e readequar o uso destes instrumentos para trazer maior resguardo ao interesse público e uma efetiva aplicação das limitações e exceções já existentes. Isto inclui:

- A possibilidade da inclusão de uma limitação própria para usos não-expressivos, tal como observado na legislação japonesa, no debate regulatório sobre Direito Autoral;
- A implementação de limitações legais às medidas de proteção tecnológica quando o fim é a efetivação de uma das limitações de Direito Autoral previstas em lei. Não seria a primeira regulação no mundo neste sentido – como já visto, o Reino Unido e até mesmo os Estados Unidos possuem disposições similares;
- No mesmo sentido, a implementação de dispositivos legais que previnam a sobreposição de limitações e exceções pela via contratual – tal como visto, novamente, no Reino Unido;
- Um maior reconhecimento do papel do Judiciário no processo de construção das limitações e exceções, por meio da cláusula aberta hoje consolidada pelo STJ – incluindo, por exemplo, uma maior conscientização – tanto para os usuários como entre juristas – sobre suas possibilidades de uso.

Por outro ângulo, o rompimento com visões polarizadas sobre a gestão do conhecimento também envolve compreender que o fomento ao processo criativo e a melhores condições de existência para os autores vai além da própria regulação estatal como um todo. Notam-se diversas questões que permanecem sem solução e que contribuem para a continuidade do atual cenário de incerteza regulatória que assola autores, pesquisadores e os demais agentes envolvidos na produção de conhecimento e cultura, as quais parecem se situar além da

capacidade de alcance do Estado. Cabe, portanto, reconhecer a necessidade de se mobilizar os entes interessados para pensar e formular alternativas à intervenção estatal em diversas áreas, tais como:

- Uma discussão ampla sobre quais modelos de remuneração seriam mais adequados aos titulares afetados por modelos de IA Generativa, com destaque especial à necessidade de se garantir uma distribuição justa e equitativa aos artistas;
- Elaboração de manuais e guias de boas práticas de pesquisa, documentação e uso de dados na pesquisa científica, bem como discutir a possibilidade de um padrão geral para planos de gestão de dados;
- Iniciativas de conscientização sobre direitos de autor na academia e incentivar a mobilização e participação da classe acadêmica na arena política e jurídica;
- Criação de modelos e repositórios abertos, capazes de atender às exigências legais de transparência e incentivar o compartilhamento de dados;
- Iniciativas de gestão de dados e obras dentro de uma lógica de “proteger para compartilhar”, visando o asseguramento da atribuição de autoria e integridade dos dados, e visando reduzir assimetrias de poder entre Norte e Sul Global, bem como entre autores e grandes titulares.

Feitas estas observações, é necessário destacar que esta pesquisa não se propõe como uma resposta definitiva ao tema da mineração de textos e dados na propriedade intelectual. Pelo contrário, há diversas questões que permanecem em aberto e não possuem resolução simples, merecendo maior aprofundamento em pesquisas futuras. A primeira delas envolve a realização de mais entrevistas dentro da comunidade acadêmica brasileira, visando o maior escopo possível quanto às regiões e áreas do conhecimento representadas, a fim de verificar, complementar e acompanhar mudanças em relação aos resultados encontrados neste trabalho.

Outra questão de suma importância consiste em explorar atentamente as diferentes formas de remuneração e controle sobre o uso de obras no ambiente digital. Afinal, a simples criação de qualquer direito exclusivo novo não tem o condão de resolver, por si só, problemas que, ao fim e ao cabo, são de ordem laboral e contratual. Ao que tudo indica, os efeitos da inteligência artificial sobre a indústria cultural adicionam dificuldades extras merecem um trabalho regulatório cuidadoso e o diálogo contínuo com artistas, pesquisadores, museus, arquivos, educadores, desenvolvedores e titulares. A 45ª sessão do SCCR na OMPI é prova da diversidade de posicionamentos e questionamentos quanto à forma pela qual os trabalhos de regulação

desta matéria devem prosseguir – e a garantia da justa compensação e proteção do trabalho criativo e acadêmico constitui parte deste problema.

Este momento de mudança também representa o maior desafio enfrentado por este trabalho: o fato de o objeto desta pesquisa ser dinâmico, em constante construção e mudança. Embora tenhamos visto que as discussões sobre uma regulação para mineração de textos e dados no Direito Autoral tenham perdurado por mais de duas décadas a nível internacional, o fato é que as tecnologias catalisadoras desse processo seguem evoluindo cada vez mais depressa – com a IA Generativa provocando questionamentos e reformulações mesmo em locais que já tinham o assunto da mineração de dados como “encaminhado”, como no caso europeu. No Brasil, o Projeto de Lei sobre IA ainda segue em tramitação, estando, logicamente, sujeito a mais mudanças – inclusive após a conclusão desta tese. Nesse sentido, se entendermos toda publicação acadêmica como uma “fotografia” de um objeto em um dado ponto no tempo, então este trabalho seria uma tentativa de fotografar um animal ou uma criança, que se recusam a parar para a câmera.

Por outro lado, constata-se, com certo cinismo, a perpetuação de um sistema de PI que sublima os interesses dos titulares e maximaliza a proteção ao mesmo tempo em que confere pouca segurança real aos autores e coloca o interesse público em posição secundária. Trata-se de um pensamento que já não é compatível com o regime constitucional brasileiro vigente, tampouco com o projeto de desenvolvimento assumido pelo país.

É preciso sedimentar a ideia de que herdar uma tradição jurídica não implica em dizer que ela está devidamente ajustada à realidade social e econômica local, e que uma postura crítica em relação à uniformização jurídica deve sempre ser mantida. Em outras palavras, o medo da precarização do trabalho não pode ceder à ilusão de que um sistema historicamente pensado por (e para) uma minoria pode ser revertido para o bem da maioria sem mudanças significativas em sua abordagem. Nesse sentido, o Brasil já demonstrou, no passado recente, a capacidade jurídica e institucional de se mover em um caminho diferente, ainda dentro das balizas acordadas em âmbito internacional. É possível e desejável se manter neste caminho, a fim de garantir que *todos* os autores – artistas e pesquisadores – saiam ganhando neste processo.

REFERÊNCIAS

Academia Brasileira de Ciências. **UNIVERSIDADES PÚBLICAS RESPONDEM POR MAIS DE 95% DA PRODUÇÃO CIENTÍFICA DO Brasil**. 15 abr. 2019. Disponível em: <https://www.abc.org.br/2019/04/15/universidades-publicas-respondem-por-mais-de-95-da-producao-cientifica-do-brasil/>. Acesso em 12 ago. 2024.

Alemanha. Act on Copyright and Related Rights (Copyright Act, as amended up to Act of September 1, 2017). Disponível em: <https://www.wipo.int/wipolex/en/text/474282>. Acesso em 22 nov. 2023.

Alvarenga, M. B. **Mineração de dados, Big Data e direitos autorais no Brasil**. 2019. Dissertação (Mestrado em Políticas Públicas, Estratégias e Desenvolvimento) – Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Andreoli-Versbach, P.; Mueller-Langer, F. Open Access to Data: An Ideal Professed but not Practised. **RatSWD Working Papers** 215, German Data Forum (RatSWD), 2013.

Antoniutti, C. L. **Usos do Big Data em Campanhas Eleitorais**. Tese de doutorado. UFRJ, ago. 2015.

Aos Fatos. **Nosso Método**. 2021. Disponível em: [https://web.archive.org/web/20210510104827/https://www.aosfatos.org/nosso-m\[%C3%A9\]todo/](https://web.archive.org/web/20210510104827/https://www.aosfatos.org/nosso-m[%C3%A9]todo/). Acesso em 30 out. 2024.

Aos Fatos. **Radar**. Disponível em: <https://www.aosfatos.org/radar/#!/>. Acesso em 30 out. 2024.

Ariente, E. A. Políticas de acesso aberto para trabalhos científicos: interesse público e direitos de autor. **Revista brasileira de políticas públicas**, vol. 7 (1), 2017.

Arnold, L.G., Bauer, C. **On the growth and welfare effects of monopolistic distortions**. *Journal of Economics*, vol. 97, 2009, pp. 19–40.

Arruda, V. C. G. **Big Data, dados pessoais e propriedade intelectual: limites e tutela jurídica das bases de dados**. 2019. 98 f. Dissertação (Direito Político e Econômico) - Universidade Presbiteriana Mackenzie, São Paulo.

Ascensão, J. O. **Direito da Internet e da sociedade da informação: estudos**. Rio de Janeiro: Forense, 2002.

Ascensão, J. O. **Direito de Autor sem Autor e Sem Obra**. Boletim da Faculdade de Direito: Separata de Estudos em Homenagem ao Professor Doutor António Castanheira Neves, vol. 2. Coimbra: Universidade de Coimbra, 2008.

Ascensão, J. O. **O Direito Autoral numa Perspectiva de Reforma**. In: Wachowicz, M.; Santos, M. J. P. (org). **Estudos de Direito de Autor: A Revisão da Lei de Direitos Autorais**. Florianópolis: Fundação Boiteux, 2010.

Ashraf, R. **Scraping EDGAR With Python**. Journal of Education for Business, 92:4, 2017, p. 179-185. Disponível em: <https://ssrn.com/abstract=3230156>. Acesso em 01 out. 2024.

Authors Guild. **Open Letter to Generative AI Leaders**. 2023a. Disponível em: <https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders>. Acesso em 02 out. 2024.

Authors Guild. **Survey Reveals 90 Percent of Writers Believe Authors Should Be Compensated for the Use of Their Books in Training Generative AI**. 2023b. Disponível em: <https://authorsguild.org/news/ai-survey-90-percent-of-writers-believe-authors-should-be-compensated-for-ai-training-use/>. Acesso em 03 out. 2024.

Associação de Jornalismo Digital. **Nota sobre o PL2338/2023 e os direitos autorais**. Brasília, 8 jul. 2024. Disponível em: <https://ajor.org.br/nota-sobre-o-pl2338-2023-e-os-direitos-autorais/>. Acesso em 26 jul. 2024.

Band, J. **Protecting User Rights Against Contract Override**. PIJIP/TLS Research Paper Series no. 97, mai. 2023. Disponível em: <https://digitalcommons.wcl.american.edu/research/97>. Acesso em 28 out. 2024.

Banterle, F. Data ownership in the data economy: a European dilemma. **EU Internet Law in the digital era (edited volume based on the REDA 2017 conference)**. Springer, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3277330. Acesso em 25 out. 2024.

Barbosa, D. B. **A noção de Originalidade e os Títulos de Obra, em particular, de Software**. 2005. Disponível em: <http://denisbarbosa.addr.com/originalidade.pdf>. Acesso em 25 out. 2024.

Barbosa, P. M. N. **Originalidade em Crise**. Revista Brasileira de Direito Civil – RBDCivil, v. 15, p. 33-48, jan./mar. 2018.

Bardin, L. **Análise de conteúdo**. Tradução de Luís Antero Reto, Augusto Pinheiro. 1ª ed. São Paulo: Edições 70, 2011.

Beiter, K. **Reforming Copyright or Toward Another Science? A More Human Rights-Oriented Approach Under the REBSPA in Constructing a "Right to Research" for Scholarly Publishing.** (2022) PIJIP/TLS Research Paper Series no. 7. Disponível em: <https://digitalcommons.wcl.american.edu/research/79>. Acesso em 28 out. 2024.

Bertón, M.J. **Text and Data Mining Exception in South America: A Way to Foster AI Development in the Region.** *GRUR International*, v. 70 (12), dez. 2021, p. 1145–1157. Disponível em: <https://doi.org/10.1093/grurint/ikab081>. Acesso em 28 set. 2024.

Bethesda Statement on Open Access Publishing. Released June 20, 2003. Disponível em: <http://legacy.earlham.edu/~peters/fos/bethesda.htm>. Acesso em 27 out. 2024.

Bittencourt, P.; Rauen, A. **Políticas de inovação: racionalidade, instrumentos e coordenação.** In: Rapini, M., Ruffoni, J.; Silva, L., Mota e Albuquerque, E. (Org) **Economia da ciência, tecnologia e inovação: fundamentos teóricos e a economia global.** 2a. edição Belo Horizonte: FACE-UFMG, 2021, Capítulo 20.

Bobbio, N. **Teoria Geral do Direito.** Traduzido por Denise Agostinetti. 3ª ed. São Paulo: Martins Fontes, 2010.

Borgman, C. **The conundrum of sharing research data.** *Journal of the American Society for Information Science and Technology* (on line), v. 63, n. 6. p. 1059-1078, 2012. Disponível em: <http://dx.doi.org/10.2139/ssrn.1869155>. Acesso em 10 jan. 2022.

Botero, C. Latin American AI Strategies Can Tackle Copyright as a Legal Risk for Researchers. **Carnegie Endowment for International Peace**, 30 abr. 2024. Disponível em: <https://carnegieendowment.org/research/2024/04/latin-american-ai-strategies-can-tackle-copyright-as-a-legal-risk-for-researchers?lang=en>. Acesso em 28 set. 2024.

Bounegru, L.; Gray, L. (eds.). **The Data Journalism Handbook: Towards a Critical Data Practice.** Amsterdam: Amsterdam University Press, 2021.

Branco, S. V. **O Domínio Público no Direito Autoral Brasileiro – Uma Obra em Domínio Público.** Rio de Janeiro: Lumen Juris, 2011.

Brasil. **Ato do Presidente do Senado Federal nº 4, de 2022.** Institui Comissão de Juristas responsável por subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei nºs 5.051, de 2019, 21, de 2020, e 872, de 2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil. 2022a. Disponível em <https://legis.senado.leg.br/atividade/comissoes/comissao/2504/mna/relatorios>. Acesso em 02 out. 2024.

Brasil. Câmara dos Deputados. **Parecer final**. 29 set 2021. Relatoria: Deputada Luisa Canziani. 2021a. Disponível em:

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2083275&file-name=Tramitacao-PL%2021/2020. Acesso em 02 out. 2024.

Brasil. Conselho da Justiça Federal. **III Jornada de Direito Comercial: Enunciados aprovados em 7/6/2019**. 2019a. Disponível em: https://www.cjf.jus.br/cjf/noticias/2019/06-junho/iii-jornada-de-direito-comercial-e-encerrada-no-cjf-com-aprovacao-de-enunciados/copy_of_EnunciadosaprovadosIIIJDCREVISADOS004.pdf. Acesso em 23 jun. 2022.

Brasil. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 28 out. 2024.

Brasil. Decreto nº 9.854, de 25 de junho de 2019. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas. 2019b. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9854.htm. Acesso em 28 out. 2024.

Brasil. Decreto nº 10.886, de 7 de dezembro de 2021. Institui a Estratégia Nacional de Propriedade Intelectual. 2021b. Brasil: Diário Oficial da União. Publicado no D.O.U. de 08 de dezembro de 2021.

Brasil. Lei nº 9.609, de 19 de fevereiro de 1998. Dispõe sobre a proteção da propriedade intelectual de programa de computador, sua comercialização no País, e dá outras providências. 1998a. Disponível em: https://www.planalto.gov.br/ccivil_03/Leis/L9609.htm. Acesso em 25 out. 2024.

Brasil. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. 1998b. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em 25 out. 2024.

Brasil. Lei nº 10.406, de 10 de janeiro de 2002. Institui o Código Civil. Disponível em: https://www.planalto.gov.br/ccivil_03/LEIS/2002/L10406compilada.htm. Acesso em 28 out. 2024.

Brasil. Lei nº 10.973, de 2 de dezembro de 2004. Dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.973.htm. Acesso em 27 out. 2024.

Brasil. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em 28 out. 2024.

Brasil. Lei nº 13.243, de 11 de janeiro de 2016. Dispõe sobre estímulos ao desenvolvimento científico, à pesquisa, à capacitação científica e tecnológica e à inovação e altera a Lei nº 10.973, de 2 de dezembro de 2004, a Lei nº 6.815, de 19 de agosto de 1980, a Lei nº 8.666, de 21 de junho de 1993, a Lei nº 12.462, de 4 de agosto de 2011, a Lei nº 8.745, de 9 de dezembro de 1993, a Lei nº 8.958, de 20 de dezembro de 1994, a Lei nº 8.010, de 29 de março de 1990, a Lei nº 8.032, de 12 de abril de 1990, e a Lei nº 12.772, de 28 de dezembro de 2012, nos termos da Emenda Constitucional nº 85, de 26 de fevereiro de 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13243.htm#art2. Acesso em 27 out. 2024.

Brasil. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 28 set. 2024.

Brasil. Ministério da Ciência, Tecnologia e Inovações. **Estratégia Brasileira de Inteligência Artificial**. 2021c. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivos/inteligenciaartificial/ebia-diagramacao_4-979_2021.pdf. Acesso em 28 set. 2024.

Brasil. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Política de Propriedade Intelectual das Instituições Científicas e Tecnológicas do Brasil – Relatório FORMICT, 2018**. Brasília – DF, 2019. 2019c.

Brasil. Ministério da Cultura. Coordenação-Geral de Regulação de Direitos Autorais. **Nota Técnica nº 2/2024, Processo nº 01400.025216/2024-13**. 2024a. Disponível em https://anpd-super.mj.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?yPDSzXhdoNcWQHJaQIHJmJIqCNXRK_Sh2SMdn1U-tzNsXmaDPgth6TWQb-VegDsJUpxKCGrIfpqgd2LPweVjfrVUs865P0Y_ahlycyNue1ge1wXB9NIS-a8VWrNBh-Xw. Acesso em 10 out. 2024.

Brasil. Ministério da Saúde. Blog da Saúde. **Sequenciamento do coronavírus possibilita o desenvolvimento de vacinas**. 16 mar. 2020. 2020a. Disponível em: <https://web.archive.org/web/20210309015721/http://www.blog.saude.gov.br/index.php/perguntas-e-respostas/54104-confira-a-entrevista-sobre-o-sequenciamento-do-coronavirus>. Acesso em 30 out. 2024.

Brasil. **REGULAMENTO DA COMISSÃO DE JURISTAS RESPONSÁVEL POR SUBSIDIAR A ELABORAÇÃO DE MINUTA DE SUBSTITUTIVO AOS PROJETOS DE LEI SOBRE INTELIGÊNCIA ARTIFICIAL**. 30 mar. 2022. 2022b. Disponível em:

<https://legis.senado.leg.br/sdleg-getter/documento/download/211b69b7-dfcf-45fd-9de6-763becafca90>. Acesso em 28 set. 2024.

Brasil. Requerimento nº 722, de 2023. Requer criação de Comissão Temporária Interna, para examinar os projetos concernentes ao relatório final aprovado pela Comissão de Juristas sobre Inteligência Artificial. 15 ago 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/159159>. Acesso em 02 out. 2024.

Brasil. Senado Federal. Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil (CJUSBIA). Relatório Final. 2022c. Disponível em: <https://legis.senado.leg.br/comissoes/mnas?codcol=2504&tp=4>. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 5.691 de 2019. Institui a Política Nacional de Inteligência Artificial, com o objetivo de estimular a formação de um ambiente favorável ao desenvolvimento de tecnologias em Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 21 de 2020. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. 2021c. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1651257291876&disposition=inline>. Acesso em 23 jun. 2022.

Brasil. Senado Federal. Projeto de Lei nº 2.338 de 2023. Dispõe sobre o uso da Inteligência Artificial. Complementação de voto. 04 jul. 2024. 2024b. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1726246478273&rendition_principal=S&disposition=inline. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 2.338 de 2023. Dispõe sobre o uso da Inteligência Artificial. Emendas 24, 44, 45, 55 e 56. 2024c. Disponíveis em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233#emendas>. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 2338 de 2023. Dispõe sobre o uso da Inteligência Artificial. Parecer do Relator. 07 jun. 2024. 2024d. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9630164&ts=1726246475112&rendition_principal=S&disposition=inline. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 2338 de 2023. Dispõe sobre o uso da Inteligência Artificial. Parecer do Relator. 18 jun. 2024. 2024e. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9640105&ts=1726246473679&rendition_principal=S&disposition=inline. Acesso em 02 out. 2024.

Brasil. Senado Federal. Projeto de Lei nº 2338 de 2023. Dispõe sobre o uso da Inteligência Artificial. Relatório Preliminar. 24 abr. 2024. 2024f. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento/download/37c068d8-46d7-472e-99bf-c3cf7afea396>. Acesso em 02 out. 2024.

Brasil. Senado Federal. Relatório aos projetos de lei que dispõem sobre a regulamentação do uso da inteligência artificial no Brasil. 07 jun. 2024. 2024g. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9630164&ts=1738768173816&rendition_principal=S&disposition=inline. Acesso em 31 jan. 2025.

Brasil. Superior Tribunal de Justiça. 3ª Turma. Recurso Especial nº 906.269/BA, relator Ministro Humberto Gomes de Barros, julgado em 16/10/2007, Lex: DJ de 29/10/2007.

Brasil. Superior Tribunal de Justiça. 3ª Turma. Recurso Especial nº 964404/ES (2007/0144450-5). Recorrente: Mitra Arquidiocesana de Vitória. Recorrido: Escritório Central de Arrecadação e Distribuição (ECAD). Relator: Min. Paulo de Tarso Sanseverino. Brasília, 15 de março de 2011. **Lex:** Diário de Justiça Eletrônico, Brasília, v. 815, 23 mai. 2011.

Brasil. Superior Tribunal de Justiça. 4ª Turma. Recurso Especial nº 1.450.302/RJ. Recorrente: Ricardo Strang. Recorrido: Access Editora Ltda. Relator: Ministro Marco Buzzi. Brasília, 11 de fevereiro de 2020. **Lex:** Diário de Justiça Eletrônico, Brasília, 14 fev. 2020. 2020b.

Brasil. Superior Tribunal de Justiça. 3ª Turma. Recurso Especial nº 1.528.627/SC, relator Ministro Paulo de Tarso Sanseverino, Terceira Turma, julgado em 7/3/2017. **Lex:** DJe de 14/3/2017.

Brasil. Supremo Tribunal Federal. Recurso Extraordinário nº 113.505-RJ. Primeira Turma. Relator: Min. Moreira Alves. Brasília, 28 de fevereiro de 1989. Disponível em: www.stf.jus.br. Acesso em 28 out. 2024.

Brasil. Supremo Tribunal Federal. Recurso Extraordinário n. 201.819 – RJ. Segunda Turma. Relatora originária: Ministra Ellen Gracie. Relator para o Acórdão: Ministro Gilmar Mendes. Brasília, 11 de outubro de 2005. Disponível em: www.stf.jus.br. Acesso em 28 out. 2024.

Braunstein, Y. M. ECONOMIC IMPACT OF DATABASE PROTECTION IN DEVELOPING COUNTRIES AND COUNTRIES IN TRANSITION. **OMPI, Standing Committee on Copyright and Related Rights, Seventh Session**, 4 abr. 2002.

Bridy, A. The Evolution of Authorship: Work Made by Code. **Colum. J.L. & Arts**, v.95, 2016.

Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. **Computer Networks and ISDN Systems**, v. 30 (1–7), abr. 1998, pp. 107-117. Disponível em: [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X). Acesso em 28 set. 2024.

Brittain, B. US judge trims AI copyright lawsuit against Meta. **Reuters**, 9 nov. 2023. Disponível em: <https://www.reuters.com/legal/litigation/us-judge-trims-ai-copyright-lawsuit-against-meta-2023-11-09/>. Acesso e 29 fev. 2024.

Brittain, B. Meta knew it used pirated books to train AI, authors say. **Reuters**, 9 jan. 2025. Disponível em: <https://www.reuters.com/technology/artificial-intelligence/meta-knew-it-used-pirated-books-train-ai-authors-say-2025-01-09/>. Acesso em 5 fev. 2025.

Brown, K. **Digital Rights Management: Trafficking in Technology That Can Be Used to Circumvent the Intellectual Property Clause**. 40 *Houston Law Review*, vol. 803, 2003, p. 803-836.

Budapest Open Access Initiative. **Budapest Open Access Initiative**. Budapeste, 14 fev. 2002. Disponível em: <https://web.archive.org/web/20240702020016/https://www.budapestopenaccessinitiative.org/read/>. Acesso em 28 set. 2024.

Bush, V. **Science: The endless frontier**. 1960. Disponível em: <https://archive.org/details/scienceendlessfr00unit/page/n9/mode/2up>. Acesso em 27 out. 2024.

Butler, S. **INSIDE THE GLOBAL DIGITAL MUSIC MARKET. OMPI, Standing Committee on Copyright and Related Rights, Forty-First Session (SCCR/41/2)**. Genebra, 1 jun. 2021.

Cabral, P. **A nova Lei de Direitos Autorais**. 4. ed. São Paulo: Harbra, 2003.

Cai, Y.; Lattu, A. Civically grounded Triple Helix: synergies between Triple Helix and Quadruple Helix models of innovation. **XVII International Triple Helix Conference**, Cape Town, 2019.

Calboli, I.; Hwang, G. **REPORT ON THE ONLINE MUSIC MARKET AND MAIN BUSINESS MODELS IN ASIA: OVERVIEW AND GENERAL TRENDS. OMPI, Standing Committee on Copyright and Related Rights, Forty-First Session (SCCR/41/7)**. Genebra, 1 jun. 2021.

Calheiros, G. C. **Estudo da Dinâmica dos Processos de Prospeção e Inovação no Ambiente das Empresas de Base Tecnológica: Case Porto Digital**. 2009. Dissertação (Mestrado em Engenharia de Produção) - 2009. Tese (Doutorado em Direito Civil) – Centro de Tecnologia e

Geociências, Programa de Pós-Graduação em Engenharia de Produção, Universidade de Pernambuco, Recife.

Carayannis, E. G.; Campbell, D. F. J. 'Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem. **International Journal of Technology Management**, v. 46, n. 3-4, p. 201- 234, 2009.

Carayannis, E. G.; Campbell, D. F. J. Open Innovation Diplomacy and a 21st Century Fractal Research, Education and Innovation (FREIE) Ecosystem: Building on the Quadruple and Quintuple Helix Innovation Concepts and the "Mode 3" Knowledge Production System. **Journal of Knowledge Economy**, v. 2, p. 327-372, 2011.

Carayannis, E. G.; Rakhmatullin, R. The Quadruple/Quintuple Innovation Helixes and Smart Specialisation Strategies for Sustainable and Inclusive Growth in Europe and Beyond. **Journal of Knowledge Economy**, v. 5, p. 212-239, 2014. Disponível em: <https://doi.org/10.1007/s13132-014-0185-8>. Acesso em 25 out. 2024.

Carboni, G. C. **Função social do direito de autor**. Curitiba: Juruá, 2006.

Carlini, N.; Hayes, J.; Nasr, M.; Jagielski, M.; Schwag, V.; Tramer, F.; Balle, B.; Ippolito, D.; Wallace, E. **Extracting Training Data from Diffusion Models**. 30 jan. 2023. Disponível em: <http://arxiv.org/abs/2301.13188>. Acesso em 29 out. 2024.

Carrier, M. **Copyright and Innovation: The Untold Story**. Wisconsin Law Review, vol. 891, out. 2012.

Carroll, M. W. **Copyright and the Progress of Science: Why Text and Data Mining is Lawful**. Washington College of Law Research Paper No. 2020-15. American University Washington College of Law, 2020.

Caspers, M.; Guibault, L. **A right to 'read' for machines: Assessing a black-box analysis exception for data mining**. Proc. Assoc. Info. Sci. Tech., 53: 1-5, 2016. Disponível em: <https://doi.org/10.1002/pr2.2016.14505301017>. Acesso em 06 jan. 2023.

Castle, C. L., Feijóo, C. STUDY ON THE ARTISTS IN THE DIGITAL MUSIC MARKET-PLACE: ECONOMIC AND LEGAL CONSIDERATIONS. **OMPI, Standing Committee on Copyright and Related Rights Forty-First Session**. Genebra, 1 jun. 2021.

Chang, H. **Is Water H2O?: Evidence, Realism and Pluralism**. New York: Springer, 2012.

Chapman, A. Towards an Understanding of the Right to Enjoy the Benefits of Scientific Progress and Its Applications. 8 **J. OF H. RTS.** 1, 2009.

Chapman, A.; Wyndham, J. A Human Right to Science. **Science**, v. 340, 14 jun. 2013.

Charquero, P. D. **Flexibilidades al derecho de autor en América Latina**. Datysoc e Fundación Karisma, abr. 2021.

Charquero, P. D.; Rangel, V. **Políticas de Inteligencia Artificial y Derechos de Autor em América Latina**. Alianza de la Sociedad Civil Latinoamericana, fev. 2023.

Chaves, A. **Direito de Autor: Princípios Fundamentais**. Rio de Janeiro: Forense, 1987.

Chen, H; Chiang, R. H. L.; Storey, V. C. **Business Intelligence and Analytics: from Big Data to Big Impact**. MIS Quarterly: Management Information Systems, vol. 36 (4), p. 1165-1188, dez. 2012.

Chen, M.; Mao, S.; Liu, Y. **Big Data: A Survey**. Networks and Applications, vol. 19 (2) New York: Springer, abr. 2014.

Chesbrough, H. **Open Innovation: The New Imperative for Creating and Profiting from Technology**. Boston: Harvard Business School Press, 2003.

Chile. Secretaria Geral de la Presidencia. **MENSAJE DE S.E. EL PRESIDENTE DE LA REPÚBLICA POR EL QUE INICIA UN PROYECTO DE LEY DE INTELIGENCIA ARTIFICIAL**. Santiago, 7 mai. 2024. Disponível em: <https://www.camara.cl/ver-Doc.aspx?prmID=17048&prmTIPO=INICIATIVA>. Acesso em 02 out. 2024.

Cho, W. Music Publishers Reach Deal With AI Giant Anthropic Over Copyrighted Song Lyrics: While the case is proceeding, Anthropic's Claude will no longer provide lyrics to songs owned by the music publishers or new lyrics based on copyrighted material. **The Hollywood Reporter**, 02 jan. 2025. Disponível em: <https://www.hollywoodreporter.com/business/business-news/anthropic-enforce-copyright-guardrails-ai-tools-1236098152/>. Acesso em 13 fev. 2025.

Clarivate. **Clarivate Terms**. 2024a. Disponível em: https://clarivate.com/wp-content/uploads/dlm_uploads/2024/04/Clarivate-Terms-v3.2-March-2024.pdf. Acesso em 27 set. 2024.

Clarivate. **How we approach Artificial Intelligence at Clarivate**. 2024b. Disponível em: <https://clarivate.com/ai/#overview>. Acesso em 27 set. 2024.

Coddington, M. Clarifying Journalism's Quantitative Turn: A Typology For Evaluating Data Journalism, Computational Journalism, And Computer-Assisted Reporting. **Digital Journalism**, v. 3 (3), p. 331–348, 2015. doi:10.1080/21670811.2014.976400.

Congost, R.; Gelman, J.; Santos, R. **PROPERTY RIGHTS IN LAND: INSTITUTIONAL INNOVATIONS, SOCIAL APPROPRIATIONS, AND PATH DEPENDENCE. KEY-NOTE**. Apresentado em: XVIth World Economic History Congress, Stellenbosch University, África do Sul, 9-13 jul. 2012.

Costigan, J. China Rules AI Firm Committed Copyright Infringement. **Forbes**, 29 fev. 2024. Disponível em: <https://www.forbes.com/sites/johannacostigan/2024/02/29/china-rules-ai-firm-committed-copyright-infringement/?sh=4881592b454d>. Acesso em 22 out. 2024.

Coreia do Sul. Copyright Act (amended up to Act No. 19597 of August 8, 2023). Disponível em <https://www.wipo.int/wipolex/en/text/588911>. Acesso em 14 dez. 2023.

Coriat, B. **From Natural-Resource Commons to knowledge Commons: Common Traits and Differences**. In: International Seminar on Property and Commons: new issues of shared access and innovation. Université de Paris, Sorbonne, França, 2013.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Ofício Circular nº 46/2024-DAV/CAPES**. Brasília, 03 de outubro de 2024.

Couldry, N.; Mejjias, U. A. **Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject**. *Television & New Media*, vol. 20 (4), p. 336-349, 2019a.

Couldry, N.; Mejjias, U. A. **Datafication**. *Internet Policy Review*, vol. 8 (4), p. 1-1-, nov. 2019. 2019b.

Craig, C. J. AI and Copyright. In: Martin-Bariteau, F.; Scassa, T (eds.). **Artificial Intelligence and the Law in Canada**. Toronto: LexisNexis Canada, 2021. Disponível em: <https://ssrn.com/abstract=3733958>. Acesso em: 05 fev. 2025.

Craig, C. J. **Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks**. 3 *Amer. Uni. Int'L L. Rev* 1, 2017.

Craig, C. J. Craig, Carys J., The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest. In: Ryan Abbott (ed.) **Research Handbook on Intellectual Property and Artificial Intelligence**. Edward Elgar, 2022. Disponível em: <https://ssrn.com/abstract=4014811> ou <http://dx.doi.org/10.2139/ssrn.4014811>. Acesso em 11 fev. 2024.

Crawford, K. **Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence**. Londres: Yale University Press, 2021.

Creswell, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 2.ed. Porto Alegre: Artmed, 2007.

Crews, K. D. **ARCHIVES COPYRIGHT EXCEPTIONS: TYPOLOGY ANALYSIS**. Organização Mundial da Propriedade Intelectual, Standing Committee on Copyright and Related Rights, Thirty-Ninth Session (SCCR/39/5). Genebra, 16 set. 2019. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_5.pdf. Acesso em 29 out. 2024.

Cruz, C. H. B. **100 universidades e institutos no Brasil com mais artigos científicos**. Fonte: **Clarivate, Incites**. 04 out. 2019. Disponível em: <https://web.archive.org/web/20240414113454/http://ciencianarua.net/wp-content/uploads/2019/04/100-univ-inst-com-mais-artigos-2014-a-2018-Brasil.pdf>. Acesso em 28 out. 2024.

Davison, M. J. **The Legal Protection of Databases**. Nova York: Cambridge University Press, 2003.

Dean, J. **Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners**. Wiley, 2014.

Declaração de Berlim sobre o Acesso Aberto ao Conhecimento nas Ciências e nas Humanidades. 2003. Disponível em: https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf. Acesso em 27 out. 2024.

Derclaye, E. **The Legal Protection of Databases: A Comparative Analysis**. Edward Elgar, 2008.

Deveci, H. A. Databases: Is Sui Generis a Stronger Bet Than Copyright? **International Journal of Law and Information Technology**, v. 12(2), p. 178-208, 2004.

Drahos, Peter; Braithwaite, John. **Information Feudalism: who owns the knowledge economy?** Nova York: The New Press, 2002.

Drexl, J.; Hilty, R. M.; Beneke, F.; Desaunettes, L. Finck, M.; Globocnik, J.; Otero, B. G.; Hoffmann, J.; Hollander, L.; Kim, D.; Richter, H.; Scheuerer, S.; Slowinski, P. R.; Thonemann, J. **Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective**. Max Planck Institute for Innovation and Competition Research

Paper Series – Research Paper No. 19-13 (Research Group on the Regulation of the Digital Economy, 2019).

Ducato, R; Strowel, A. Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility. **CRIDES Working Paper Series**, 2018. Disponível em: <https://ssrn.com/abstract=3278901> ou <http://dx.doi.org/10.2139/ssrn.3278901>. Acesso em 30 out. 2024.

Duro, L. D. **New Text and Data Mining Exceptions: Implementation in the EU and the possible outcome of Brazil adopting a similar approach**. 2019. Dissertação (Master in Intellectual Property and Knowledge Management) – Faculty of Law, Maastricht University, Maastricht, Holanda.

Dusollier, S. **Les nouvelles dispositions belges en matière de protection technique du droit d'auteur et des droits voisins**. 2005.

Dusollier, S. L'introuvable interface entre exceptions au droit d'auteur et mesures techniques de protection. **Communication Commerce Electronique**, 11, 21-24, 2006.

Eicher, T., Penalosa, C. G. **Endogenous strength of intellectual property rights: implications for economic development and growth**. *European Economic Review*, vol. 52 (2), 2008, pp. 237–258.

Eldan, R.; Russinovich, M. **Who's Harry Potter? Approximate Unlearning in LLMs 2–3, 6–8 (Oct. 4, 2023)** (unpublished manuscript). Disponível em: <https://arxiv.org/abs/2310.02238>. Acesso em 21 out. 2024.

Elsevier. **Terms and Conditions**. Disponível em: <https://www.elsevier.com/legal/elsevier-website-terms-and-conditions>. Acesso em 27 set. 2024.

Equador. **Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación. Año IV – N° 899**. Quito, 9 de dezembro de 2016. Disponível em: <https://wipo-lex.wipo.int/en/text/439750>. Acesso em 23 jun. 2022.

Estados Unidos. 17 U.S. Code § 107 - Limitations on exclusive rights: Fair use. 1976. Disponível em: <https://www.law.cornell.edu/uscode/text/17/107>. Acesso em 28 out. 2024.

Estados Unidos. 17 U.S. Code § 1201 - Circumvention of copyright protection systems. 1998. Disponível em: <https://www.law.cornell.edu/uscode/text/17/1201>. Acesso em 28 out. 2024.

Estados Unidos. House of Representatives. Copyright Law Revision. Report No. 94-1476. 3 set. 1976. 1976b. Disponível em: https://www.copyright.gov/history/law/clrev_94-1476.pdf. Acesso em 28 set. 2024.

Estados Unidos. OBJECTIVES AND PRINCIPLES FOR EXCEPTIONS AND LIMITATIONS FOR LIBRARIES AND ARCHIVES. **OMPI, Standing Committee on Copyright and Related Rights, Forty-Fourth Session**, 2 nov. 2023. 2023a. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_44/sccr_44_5.pdf. Acesso em 29 out. 2024.

Estados Unidos. Congresso dos Estados Unidos. **S.4674 - Content Origin Protection and Integrity from Edited and Deepfaked Media Act of 2024 (Bill)**. 2024a. Disponível em: <https://www.congress.gov/bill/118th-congress/senate-bill/4674?q=%7B%22search%22%3A%22content+origin+protection%22%7D&s=2&r=1>. Acesso em 30 out. 2024.

Estados Unidos. Copyright Office. **Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies**. Federal Register/Vol. 89, No. 208/Monday, October 28, 2024/Rules and Regulations. 2024b. Disponível em: <https://www.govinfo.gov/content/pkg/FR-2021-10-28/pdf/2021-23311.pdf>. Acesso em 31 mar. 2025.

Estados Unidos. Copyright Office. **Rulemaking Proceedings Under Section 1201 of Title 17**. Disponível em: <https://www.copyright.gov/1201/>. Acesso em 31 mar. 2025.

Estados Unidos. National Institute of Health. **PubMed**. 2024c. Disponível em: <https://pubmed.ncbi.nlm.nih.gov>. Acesso em 28 out. 2024.

Estados Unidos. Suprema Corte dos EUA. *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 922 (2d Cir. 1994). 1994a.

Estados Unidos. Suprema Corte dos EUA. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 208-211 (2d Cir. 2015).

Estados Unidos. Suprema Corte dos EUA. *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

Estados Unidos. Suprema Corte dos EUA. *Campbell v. Acuff-Rose Music, Inc.* 510 U.S. 569, 579, 114 S. Ct. 1164, 1171, 127 L. Ed. 2d 500 (1994). 1994b. Disponível em: <https://supreme.justia.com/cases/federal/us/510/569/>. Acesso em 28 set. 2024.

Estados Unidos. Suprema Corte dos EUA. *Harper & Row v. Nation Enterprises*, 471 U.S. 539 (1985).

Estados Unidos. Suprema Corte dos EUA. Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003).

Estados Unidos. Suprema Corte dos EUA. Leadsinger, Inc. v. BMG Music Publishing, 512 F.3d 522 (9th Cir. 2008). Disponível em: <https://casetext.com/case/leadsinger-v-bmg-music-pub>. Acesso em 28 set. 2024.

Estados Unidos. Suprema Corte dos EUA. Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).

Estados Unidos. Suprema Corte dos EUA. Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984). Disponível em: <https://supreme.justia.com/cases/federal/us/464/417/>. Acesso em 28 set. 2024.

Estados Unidos. Thomson Reuters Enterprise Centre GmbH et al v. ROSS Intelligence Inc., COMPLAINT filed with Jury Demand against ROSS Intelligence Inc. - Magistrate Consent Notice to Pltf. (Filing fee \$ 400, receipt number ADEDC-2919270.) - filed by West Publishing Corporation, Thomson Reuters Enterprise Centre GmbH. (Attachments: #1 Exhibit A, #2 Civil Cover Sheet)(mal). 2020, p. 4. Disponível em: <https://docs.justia.com/cases/federal/district-courts/delaware/dedce/1:2020cv00613/72109/1>. Acesso em 21 ago. 2024.

Estados Unidos. U.S. District Court for the District of Delaware. Getty Images (US) Inc. v. Stability AI Inc, U.S. District Court for the District of Delaware, No. 1:23-cv-00135, 03 fev. 2023. 2023b. Disponível em: <https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.1.0.pdf>. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the District of Massachusetts. UMG RECORDINGS, INC., CAPITOL RECORDS, LLC, SONY MUSIC ENTERTAINMENT, ATLANTIC RECORDING CORPORATION, ATLANTIC RECORDS GROUP LLC, RHINO ENTERTAINMENT LLC, THE ALL BLACKS U.S.A., INC., WARNER MUSIC INTERNATIONAL SERVICES LIMITED, and WARNER RECORDS INC. v. SUNO, INC. and JOHN DOES 1-10. Case 1:24-cv-11611, 24 jun. 2024. 2024d. Disponível em: <https://www.riaa.com/wp-content/uploads/2024/06/Suno-complaint-file-stamped20.pdf>. Acesso em 30 out. 2024.

Estados Unidos. U.S. District Court for the Northern District of California. Andersen v. Stability AI, U.S. District Court for the Northern District of California, No. 3:23-cv-00201, 03 jan. 2023. 2023c. Disponível em: <https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.1.0.pdf>. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the Northern District of California. Andersen v. Stability AI, U.S. District Court for the Northern District of California, No. 3:23-cv-00201, 03 jan. 2023. Order on Motions to Dismiss and Strike. 23-cv-00201-WHO, (N.D. Cal. Oct. 30, 2023).

2023d. Disponível em: <https://casetext.com/case/andersen-v-stability-ai-ltd>. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the Northern District of California. Silverman v. OpenAI Inc, U.S. District Court for the Northern District of California, No. 3:23-cv-03416, 07 jul. 2023. 2023e. Disponível em: https://storage.courtlistener.com/recap/gov.uscourts.cand.415174/gov.uscourts.cand.415174.1.0_2.pdf. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the Northern District of California. Tremblay v. OpenAI Inc, U.S. District Court for the Northern District of California, No. 3:23-cv-03223. 2024e. Disponível em: <https://casetext.com/case/tremblay-v-openai-inc-9>. Acesso em 30 out. 2024.

Estados Unidos. U.S. District Court for the Southern District of New York. New York Times Co v Microsoft Corp et al, U.S. District Court, Southern District of New York, No. 23-11195, 27 dez. 2023. 2023f. Disponível em: <https://www.courthousenews.com/wp-content/uploads/2023/12/new-york-times-microsoft-open-ai-complaint.pdf>. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the Southern District of New York. Raw Story Media v. OpenAI Inc, U.S. District Court for the Southern District of New York, No. 1:24-cv-01514, 28 fev. 2024. 2024f. Disponível em: <https://www.loevy.com/wp-content/uploads/2024/02/Raw-Story-v.-OpenAI-Complaint-Filed.pdf>. Acesso em 29 fev. 2024

Estados Unidos. U.S. District Court for the Southern District of New York. The Intercept Media Inc v. OpenAI Inc, U.S. District Court for the Southern District of New York, No. 1:24-cv-01515, 28 fev. 2024. 2024g. Disponível em: <https://www.loevy.com/wp-content/uploads/2024/02/Intercept-v.-OpenAI-Complaint-Filed.pdf>. Acesso em 29 fev. 2024.

Estados Unidos. U.S. District Court for the Southern District of New York. UMG RECORDINGS, INC., CAPITOL RECORDS, LLC, SONY MUSIC ENTERTAINMENT, ARISTA MUSIC, ARISTA RECORDS LLC, ATLANTIC RECORDING CORPORATION, RHINO ENTERTAINMENT COMPANY, WARNER MUSIC INC., WARNER MUSIC INTERNATIONAL SERVICES LIMITED, WARNER RECORDS INC., WARNER RECORDS LLC, and WARNER RECORDS/SIRE VENTURES LLC v. UNCHARTED LABS, INC., d/b/a Udio.com, and JOHN DOES 1-10. Case 1:24-cv-04777, 24 jun. 2024. 2024h. Disponível em: <https://www.riaa.com/wp-content/uploads/2024/06/Udio-Complaint-6.24.241.pdf>. Acesso em 30 out. 2024.

Estados Unidos. U.S. District Court for the Southern District of New York. Universal City Studios v. Reimerdes, 111 F. Supp. 2d 294, (S.D.N.Y. 2000). Disponível em: <https://casetext.com/case/universal-city-studios-v-reimerdes-2/>. Acesso em 30 out. 2024.

Etzkowitz, H. **Entrepreneurial Scientists and Entrepreneurial Universities in American Academic Science**. *Minerva*, v. 21, n. 2/3, p. 198-233, 1983. Disponível em:

<http://www.jstor.org/stable/41820527.pdf>. DOI: <https://doi.org/10.1007/BF01097964>. Acesso em 27 out. 2024.

Etzkowitz, H.; Leydesdorff, L. The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university-industry-government relations. **Research Policy**, v. 29, n. 2, p. 109-123, 2000.

Etzkowitz, H.; Leydesdorff, L. The Triple Helix—University-Industry-Government relations: A laboratory for knowledge based economic development. **EASST Review**, v. 14, p. 14-19, 1995.

European Copyright Society. **General Opinion on the EU Copyright Reform Package**. 24 jan. 2017. Disponível em: <https://europeancopyrightsocietydotorg.files.wordpress.com/2015/12/ecs-opinion-on-eu-copyright-reform-def.pdf>. Acesso em 04 out. 2024.

European Guild for Artificial Intelligence Regulation. **Manifesto for AI Companies Regulation in Europe**. 2023. Disponível em: <https://www.egair.eu/#manifesto>. Acesso em 03 out. 2024.

Farinha, L.; Ferreira, J. J. Triangulation of the triple helix: a conceptual framework. **Triple Helix Association**, v. 1, 2013. DOI 10.13140/2.1.4161.1202

Ferraz, J.C.; de Paula, G.; Kupfer, D. **Política industrial**. In: Kupfer, D., Hasenclever, L. **Economia industrial: fundamentos teóricos e práticas no Brasil**. 3a. ed. São Paulo: Atlas, 2002, Capítulo 24.

Fischer, B.; Schaeffer, P. R.; Silveira, J. P. Universities’ gravitational effects on the location of knowledge-intensive investments in Brazil. **Science and Public Policy**, v. 45, n. 5, p. 692-707, 2018. DOI: <https://doi.org/10.1093/scipol/scy002>.

Fischer, B. B.; Schaeffer, P. R.; Vonortas, N. S. O Papel das Universidades na Dinâmica Tecnológica Brasileira. In: Guimarães, P. B. V.; et al. **Empreendedorismo, Formação e Inovação Tecnológica em Instituições de Ensino Superior**. Natal: Editora Motres, p. 15-40, 2020.

Flick, U. **Introdução à pesquisa qualitativa**. Traduzido por Joice Elias Costa. 3ª ed. Porto Alegre: Artmed, 2009.

Flynn, S.; Geiger, C.; Quitnais, J. P.; Margoni, T.; Sag, M.; Guibault, L.; Carroll, M. W. **Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action**. PIJIP/TLS Research Paper Series, 48, 2020.

Flynn, S.; Palmedo, M. **The User Rights Database: Measuring the Impact of Copyright Balance**. 4 dez. 2017. Disponível em: <https://ssrn.com/abstract=3082371>. Acesso em 23 jun. 2022.

Flynn, S.; Palmedo, M.; Izquierdo, A. **Research Exceptions in Comparative Copyright Law**. PIJIP/TLS Research Paper Series no. 72, 2021.

Flynn, S.; Schirru, L.; Palmedo, M.; Izquierdo, A. **Research Exceptions in Comparative Copyright**. 2022a. PIJIP/TLS Research Paper Series no. 75, 2022.

Flynn, S.; Butler, B.; Carroll, M.; Cohen-Sasson, O.; Craig, C.; Guibault, L.; Jaszi, P.; Jütte, B. J.; Katz, A.; Quintais, J. P.; Margoni, T.; Souza, A. R.; Sag, M.; Samberg, R.; Schirru, L.; Senftleben, M.; Tur-Sinai, O.; Contreras, J. **Legal reform to enhance global text and data mining research: Outdated copyright laws around the world hinder research**. Science, dez. 2022b. Disponível em: <https://www.science.org/doi/epdf/10.1126/science.add6124>. Acesso em 06 jan. 2023.

Flynn, S.; Vyas, L. **Examples of text and data mining research using copyrighted materials**. InfoJustice Blog, 5 dez. 2022. Disponível em: <https://infojustice.org/archives/44948>. Acesso em 2 fev. 2023.

Foster, D. **Generative Deep Learning: Teaching Machines to Paint, Write, Compose and Pay**. Sebastopol, CA: O'Reilly, 2019.

França. Intellectual Property Code (consolidated version of April 25, 2016). Disponível em: <https://www.wipo.int/wipolex/en/legislation/details/16057>. Acesso em 29 out. 2024.

Franco, M. L. P. B. **Análise de conteúdo**. 2ª ed. Brasília: Plano, 2005.

Fundação Oswaldo Cruz. **Plano de Dados Abertos da Fundação Oswaldo Cruz (PDA Fiocruz 2023-2025)**. 2023. Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos_2/pda_fiocruz_2023_-_2025_final.pdf. Acesso em 28 out. 2024.

Furukawa, Y. **The protection of intellectual property rights and endogenous growth: Is stronger always better?** Journal of Economic Dynamics & Control, vol. 31, pp. 3644–3670, 27 fev. 2007.

Galvão, M.C.B.; Ricarte, I.L.M. **REVISÃO SISTEMÁTICA DA LITERATURA: CONCEITUAÇÃO, PRODUÇÃO E PUBLICAÇÃO**. LOGEION: Filosofia da informação, Rio de Janeiro, v. 6 n. 1, p.57-73, set.2019/fev. 2020.

Gangopadhyay, K.; Mondal, D. **Does stronger protection of intellectual property stimulate innovation?** *Economics Letters*, vol. 116, pp. 80-82, 24 jan. 2012.

Garvey, E. G., **'facts and FACTS': Abolitionists' Database Innovations**. In: Gitelman, L. **Raw Data is an Oxymoron**. The MIT Press, 2013.

Gaskell, G. **Entrevistas individuais e grupais**. In: Bauer, M.W.; Gaskell, G. (Eds.). **Pesquisa qualitativa com texto, imagem e som: um manual prático**. Petrópolis, RJ: Vozes, 2002, cap.3.

Gassmann, O.; Widenmayer, B; Friesike, S; and Schildhauer, T. **Opening Up Science: Towards an Agenda of Open Science in Industry and Academia**. HIIG Discussion Paper Series No. 2012-04, 25 out. 2011. Disponível em: <https://ssrn.com/abstract=2091122>. Acesso em 28 out. 2024.

Geiger, C. **Building an ethical framework for intellectual property in the EU: time to revise the Charter of Fundamental Rights**. In: Ghidini, G.; Falce, V. (eds.). **Reforming Intellectual Property Law**. Edward Elgar, 2022, p. 90-91.

Geiger, C. **The social function of intellectual property rights, or how ethics can influence the shape and use of IP law**. In: Dinwoodie, G. B. (ed.). **Methods and Perspectives in Intellectual Property**. Edward Elgar, 2014, p. 153-176.

Geiger, C. *et al.* **Declaration "A Balanced Interpretation Of The 'Three-Step Test' In Copyright Law"**. JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law, 2010, p. 199-122. Disponível em: <https://www.jipitec.eu/issues/jipitec-1-2-2010/2621/Declaration-Balanced-Interpretation-Of-The-Three-Step-Test.pdf>. Acesso em 25 out. 2024.

Geiger, C.; Gervais, D.; Senftleben M. **Understanding the "three-step test"**. In: Gervais, D. (Ed.), **International Intellectual Property – A Handbook of Contemporary Research**. Cheltenham: Edward Elgar, 2015, p. 167-189.

Geiger, C.; Iaia, V. **The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI (October 6, 2023)**. *Computer Law & Security Review*, vol 52 (no prelo, 2024). Disponível em: <https://ssrn.com/abstract=4594873>. Acesso em 03 out. 2024.

Geiger, C.; Izyumenko, E. **Towards a European "Fair Use" Grounded in Freedom of Expression**. *American University International Law Review*, vol. 35 (1), 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3379531. Acesso em 25 out. 2024.

Geiger, C.; Frosio, G.; Bulayenko, O. **The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (March 2, 2018)**. Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02, Disponível em: <https://ssrn.com/abstract=3160586> or <http://dx.doi.org/10.2139/ssrn.3160586>. Acesso em 23 jun. 2022.

Gervais, D. J. **Exploring the Interfaces Between Big Data and Intellectual Property Law**. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, vol. 10 (1), pp. 22-38, 2019.

GISAID. **About us. History**. Disponível em: <https://www.gisaid.org/about-us/history/>. 2024a. Acesso em 25 out. 2024.

GISAID. **Terms of Use**. Disponível em: <https://www.gisaid.org/registration/terms-of-use/>. 2024b. Acesso em 25 out. 2024.

Gonçalves, L. R. **A TUTELA JURÍDICA DE TRABALHOS CRIATIVOS FEITOS POR APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL NO BRASIL**. 2019. Dissertação (Direito) – Faculdade de Direito, Universidade Federal do Paraná (UFPR), Curitiba.

Granstrand, O. **Innovation and Intellectual Property Rights**. In: Fagerberg, J., Mowery, D. & Nelson, R. (eds.). **The Oxford Handbook of Innovation**. Oxford University Press, 2007.

Gray, C. **More than Extraction: Rethinking Data's Colonial Political Economy**. *International Political Sociology* 0, 1–20, 2023.

Griffiths, J., Synodinou, T., Xalabarder, R. **Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market**. European Copyright Society, 3 mai. 2022. Disponível em: https://europeancopyrightsociety.org/wp-content/uploads/2022/05/ecs_exceptions_final-3.pdf. Acesso em 4 set. 2024.

Grossi, P. **História da propriedade e outros ensaios**. Traduzido por Luiz Ernani Fritoli e Ricardo Marcelo Fonseca. Rio de Janeiro: Renovar, 2006.

Guadamuz, A. **A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs**. *GRUR International* 2/2024 (forthcoming 2024). Disponível em: <https://ssrn.com/abstract=4371204> ou <http://dx.doi.org/10.2139/ssrn.4371204>. Acesso em 30 out. 2024.

Guadamuz, A. **French lawmakers propose new copyright law about generative AI.** TechnoLlama, 24 set. 2023. Disponível em: <https://www.technollama.co.uk/french-lawmakers-propose-new-copyright-law-about-generative-ai>. Acesso em 03 out. 2024.

Guanaes, P. C. V. **Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral.** 2020. 234f. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Guanaes, P. C. V.; Albagli, S. Dados de pesquisa subjacentes a artigos científicos: questões do direito autoral. **Em Questão**, 2022, Vol.28 (3), p.114171-25.

Guerrero, M.; Urbano, D.; Fayolle, A.; Klofsten, M.; Mian, S. Entrepreneurial universities: emerging models in the new social and economic landscape. **Small business economics**, v. 47, n. 3, p. 551-563, 2016. DOI: <https://doi.org/10.1007/s11187-016-9755-4>.

Han, J.; Pei, J.; Kamber, M. **Data mining: concepts and techniques.** [s.l.] Elsevier, 2011.

Hardin, G. The Tragedy of the Commons. **Science**, New Series, Vol. 162, No. 3859, Dez. 13, 1968.

Have I Been Trained? 2024. Disponível em: <https://haveibeen trained.com/>. Acesso em 28 ago. 2024.

Haykin, S. **Redes neurais: princípios e práticas.** Traduzido por Paulo Martins Engel. – 2.ed. – Porto Alegre: Bookman, 2001.

Henry, K. **Importance of Web Scraping in E-Commerce and E-Marketing.** 19 jan. 2021. Disponível em: <https://ssrn.com/abstract=3769593> ou <http://dx.doi.org/10.2139/ssrn.3769593>. Acesso em 01 out. 2024.

Heravi, B. R. 3Ws of Data Journalism Education: What, Where and who? **Journalism Practice** v. 13 (3), p. 349–366, 2019. doi:10.1080/17512786.2018.1463167.

Hespanha, A. M. **O jurista e o legislador na construção da propriedade burguesa liberal em Portugal.** In *Análise Social*, vol. XVI (61-62), 1980.

Hess, C. **Mapping the Commons.** In: 12th Biennial Conference of the International Association for the Study of the Commons (IASC), University of Gloucestershire, Cheltenham, England, 2008.

Hrynaszkiewicz, I.; Birukou, A.; Astell, M.; Swaminathan, S.; Kenall, A.; Khodiyar, V. **Standardising and Harmonising Research Data Policy in Scholarly Publishing**. International Digital Curation Conference, v. 12, n. 1, p. 65-71, 2017. Disponível em: <http://www.ijdc.net/article/view/12.1.65>. Acesso em: 23 jun. 2022.

Hugenholtz, P. B. **Data property: Unwelcome Guest in the house of IP**. In: REDA, F. (ed.). **Better Regulation for Copyright: Academics meet Policy Makers**. TheGreens/EFA, p. 65-77, 2017. Disponível em: https://juliareda.eu/wp-content/uploads/2017/09/2017-09-06_Better-Regulation-for-Copyright-Academics-meet-Policy-Makers_Proceedings.pdf. Acesso em 23 jun. 2022.

IBM. **What is generative AI?** 2024c. Disponível em: <https://research.ibm.com/blog/what-is-generative-ai>. Acesso em 29 out. 2024.

Initiative Urheberrecht. **Joint Statement: Authors and Performers Call for Safeguards Around Generative AI in the European AI Act**. 19 abr. 2023. Disponível em: <https://urheber.info/diskurs/call-forsafeguards-around-generative-ai>. Acesso em 03 out. 2024.

Instituto Brasileiro de Geografia e Estatística. **Pesquisa de Inovação Semestral - PINTEC Semestral 2022**. 2024. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv102067.pdf>. Acesso em 19 jul. 2024.

Internet Archive. **Wayback Machine**. Disponível em: <https://web.archive.org>. Acesso em 23 out. 2024.

Jain, S.; George, G.; Maltarich, M. Academics or entrepreneurs? Investigating role identity modification of university scientists involved in commercialization activity. **Research Policy**, v. 38, p. 922-935, 2009. Disponível em: <https://doi.org/10.1016/j.respol.2009.02.007>. Acesso em 27 out. 2024.

Japão. Agency for Cultural Affairs. **AIと著作権**. 2020. Disponível em: https://www.bunka.go.jp/seisaku/chosakuken/pdf/93903601_01.pdf. Acesso em 30 out. 2024.

Japão. **Digital Net Jidai niokeru Chitekizaisan Senmoncho^ sakai, Digital Net Jidai niokeru Chizaiseido no Arikata nitsuite (Report) [Intellectual Property System for the Digital-Network Age]** 11. 2008.

Japão. **Bunka-Shingikai Chosakuken-Bunkakai Ho^kokusho 2017 [The Report of the Subdivision on Copyright of the Council for Cultural Affairs 2017]**. 2017a. Disponível em: https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2904_shingi_hokokusho.pdf. Acesso em 01 out. 2024.

Japão. Intellectual Property Strategy Headquarters. **Intellectual Property Strategic Program 2017**. 2017b. Disponível em: https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizai-keikaku20170516_e.pdf. Acesso em 25 out. 2024.

Japão. IoT Acceleration Consortium. **Resumo do GT da Promoção da Distribuição de Dados**. 22 jan. 2016. Disponível em: <http://www.iotac.jp/wg/data/>. Acesso em 26 out. 2024.

Japão. Ministério da Economia, Comércio e Indústria. **Information Economy**. Disponível em: https://www.meti.go.jp/english/policy/mono_info_service/information_economy/index.html. Acesso em 25 out. 2024.

Japão. Ministério da Economia, Comércio e Indústria. **Japan-EU Joint Press Statement on Facilitating the Free Flow of Data Issued**. Mar 2017. 2017b. Disponível em: https://www.meti.go.jp/english/press/2017/0321_002.html. Acesso em 25 out. 2024.

Japão. Ministério da Justiça. **Lei nº 38, de 6 de maio de 1970. Emenda nº 73 de 2009**. Disponível em: <http://www.japaneselawtranslation.go.jp/law/detail/?printID=&ft=1&re=02&dn=1&x=0&y=0&co=01&ia=03&ky=copyright+act&page=24&vm=02&lvm=02&id=3379>. Acesso em 23 jun. 2022.

Japão. Ministério da Justiça. **Lei nº 38, de 6 de maio de 1970. Emenda nº 30 de 2018**. Disponível em: <http://www.japaneselawtranslation.go.jp/law/detail/?printID=&ft=1&re=02&dn=1&x=0&y=0&co=01&ia=03&ky=copyright+act&page=24&vm=02&lvm=02&id=3379>. Acesso em 23 jun. 2022.

Jesus, J. G., Sacchi, C., Candido, D. D. S., Claro, I. M., Sales, F. C. S., Manuli, E. R., Silva, D. B. B. D., Paiva, T. M., Pinho, M. A. B., Santos, K. C. O., Hill, S. C., Aguiar, R. S., Romero, F., Santos, F. C. P. D., Gonçalves, C. R., Timenetsky, M. D. C., Quick, J., Croda, J. H. R., Oliveira, W., Rambaut, A., ... Faria, N. R. (2020). Importation and early local transmission of COVID-19 in Brazil, 2020. *Revista do Instituto de Medicina Tropical de Sao Paulo*, 62, e30. Disponível em: <https://doi.org/10.1590/s1678-9946202062030>. Acesso em 29 out. 2024.

Julien, P. A. **Empreendedorismo Regional e a Economia do Conhecimento**. São Paulo: Nova Fronteira, 2010. Cap.1.

Kant, I. **A Metafísica dos Costumes**. Traduzido por Edson Bini. 1ª ed. Bauru: EDIPRO, 2003.

Kaplan, A.; Haenlein, M. **Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence**. *Business Horizons*, vol. 62, pp. 15-25, jan-fev. 2019.

Kelleher, J. D.; Tierney, B. **Data Science**. Cambridge: MIT Press, 2018.

Keller, P. **Considerations for Opt-Out Compliance Policies by AI Model Developers**. Open Future Policy Brief #6, 16 mai. 2024. 2024a. Disponível em: https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf. Acesso em 01 out. 2024.

Keller, P. **Machine readable or not? – notes on the hearing in LAION e.v. vs Kneschke**. Kluwer Copyright Blog, 22 jul. 2024. 2024b. Disponível em: <https://copyrightblog.kluweriplaw.com/2024/07/22/machine-readable-or-not-notes-on-the-hearing-in-laion-e-v-vs-kneschke/>. Acesso em 01 out. 2024.

Knibbs, K. Every AI Copyright Lawsuit in the US, Visualized: WIRED is following every copyright battle involving the AI industry—and we’ve created some handy visualizations that will be updated as the cases progress. **WIRED**, 19 dez. 2024. Disponível em: <https://www.wired.com/story/ai-copyright-case-tracker/>. Acesso em 13 fev. 2025.

Kroenke, D. M.; Auer, D. J.; Vandenberg, S. L.; Yoder, R. C. **Database Concepts**. 8ª ed. Nova York: Pearson, 2016.

Kwan, Y.K., Lai, E.L.-C. **Intellectual property rights protection and endogenous economic growth**. *Journal of Economic Dynamics and Control* 27 (5), 2003, pp. 853–873.

Lana, P. P. **A AUTORIA DAS OBRAS AUTONOMAMENTE GERADAS POR INTELIGÊNCIA ARTIFICIAL E O DOMÍNIO DO PÚBLICO**. 2020. Dissertação (Direito, Área de Especialização em Ciências Jurídico-Empresariais / Menção em Empresarial) – Faculdade de Direito, Universidade de Coimbra, Coimbra.

Lemley, M. A. **Faith-Based Intellectual Property**. *UCLA L. REV.* 62, p. 1328, 2015.

Lessig, L. **Code and Other Laws of Cyberspace**. Nova York: Basic Books, 1999.

Lessig, L. **Cultura Livre: Como a mídia usa a tecnologia e a lei para barrar a criação cultural e controlar a criatividade**. Traduzido por Fábio Emílio Costa. 2004. Disponível em: <https://www.ufmg.br/proex/cpinfo/educacao/docs/10d.pdf>. Acesso em 30 out. 2024.

Levendowski, A. **How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem**. *Wash L. Rev.* 579, 2018.

Lewicki, B. C. **Limitações aos direitos de autor: releitura na perspectiva do direito civil contemporâneo**. 2007. 299 f. Tese (Doutorado em Direito Civil). Centro de Ciências Sociais, Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro.

Lin, J.; Strasser, C. **Recommendations for the Role of Publishers in Access to Data**. *PLoS Biology*, v. 12, n. 10, e1001975, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4211645>. Acesso em: 10 jan. 2022.

Liu, S.; van der Sijde, P. C. Towards the Entrepreneurial University 2.0: Reaffirming the Responsibility of Universities in the Era of Accountability. **Sustainability**, v. 13 (6), p. 3073, 2021. Disponível em: <https://doi.org/10.3390/su13063073>. Acesso em 27 out. 2024.

Lobato, Luisa Cruz. **A world made of apps? Algorithms and (in)security governance in the Global South**. Rio de Janeiro, 2021. 277p. Tese (Doutorado em Relações Internacionais) – Instituto de Relações Internacionais, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Locke, J. **Second Treatise of Government**. 1764. Disponível em: <https://www.gutenberg.org/files/7370/7370-h/7370-h.htm>. Acesso em 25 out. 2024.

López, A. EL IMPACTO DE LA PROTECCIÓN DE LAS BASES DE DATOS NO ORIGINALES SOBRE LOS PAÍSES DE AMÉRICA LATINA Y EL CARIBE. **OMPI, Comité Permanente de Derecho de Autor y Derechos Conexos, Octava Sesión**, 15 out. 2002.

Luger, G. F. **Inteligência Artificial: Estruturas e estratégias para a resolução de problemas complexos**. 4ª ed. Traduzido por Paulo Martins Engel. Bookman (Artemed)/Pearson Education Limited 2002. 2004.

Lunney, G. S. **The Death of Copyright: Digital Technology, Private Copying, and the DMCA**. Disponível em: <https://ssrn.com/abstract=263208>. Acesso em 09 out. 2024.

Machado, J. **Dados abertos e ciência aberta**. In: Albagli, S.; Maciel, M. L.; Abdo, A. H.(eds.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015, p. 201-228.

Malerba, F. SECTORAL SYSTEMS OF INNOVATION AND PRODUCTION. **RESEARCH POLICY**, v. 31 (2), fev. 2002. Disponível em: https://www.ige.unicamp.br/pos-graduacao/wp-content/uploads/2015/30/shared-media/Potential%20Proposals/Groups/Group%202_Current%20challenges%20of%20renewable%20energies%20and%20biopharma_a%20comparative%20innovation%20systems%20study/malerba_sectoral2002.pdf. Acesso em 01 mai 2025.

Manso, E. V. **Os fundamentos da exclusão de ilicitude de atos que contrariam direitos autorais.** Revista dos Tribunais. n. 557, mar. 1982, pp. 255-260.

Margoni, T. Saving research: Lawful access to unlawful sources under Art. 3 CDSM Directive? **Kluwer Copyright Blog**, 22 dez. 2023. Disponível em: <https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/>. Acesso em 4 set. 2024.

Marques, J. Usando métodos computacionais na pesquisa em acervos digitais. In: Núcleo de Informação e Coordenação do Ponto BR; Comitê Gestor da Internet no Brasil. **TIC CULTURA: Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Equipamentos Culturais Brasileiros.** São Paulo: 2023.

Martens, B. **The importance of data access regimes for artificial intelligence and machine learning.** Digital Economy Working Paper 2018-09, JRC Technical Reports. Sevilha: Comissão Europeia, 2018.

Marzano, F. M. **Políticas de inovação no Brasil e nos Estados Unidos: a busca da competitividade – oportunidades para a ação diplomática.** Brasília: Fundação Alexandre Gusmão, 2011.

Mazzucato, M. **The Entrepreneurial State: Debunking Public vs. Private Sector Myths.** Londres e Nova York: Anthem Press, 2013.

Mcafee, A.; Brynjolfsson, E. **Big Data: The Management Revolution.** Harvard Business Review, pp. 1-9, out. 2012.

Merton, R.K. **Priorities in scientific discovery: a chapter in the sociology of Science.** American Sociological Review, 22(6), 635-659, 1957.

Michel, J. Direito de autor, direito de cópia e direito à informação: o ponto de vista e a ação das associações de profissionais da informação e da documentação. **Ciência da informação**, 1997, Vol.26 (2).

Mitchell, M. **Complexity: a guided tour.** Oxford University Press, 2009.

Mueller-Langer, F.; Scheufen, M. Academic Publishing and Open Access. In: Handke, C.; Towse, R. (eds.), **Handbook of the Digital Creative Economy**, Edward Elgar, 2013, pp. 365-377, Max Planck Institute for Intellectual Property & Competition Law Research Paper No. 13-03. Disponível em: <https://ssrn.com/abstract=2198400>. Acesso em 27 out. 2024.

Murray, F.; Stern, S. **Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis.** *Journal of Economic Behavior & Organization*, vol. 63, pp. 648–687, 16 jan. 2007.

Murthy, D., Gross, A., Takata, A., Bond, S. **Evaluation and Development of Data Mining Tools for Social Network Analysis.** In: Özyer, T., Erdem, Z., Rokne, J., Khoury, S. (eds) **Mining Social Networks and Security Informatics. Lecture Notes in Social Networks.** Dordrecht: Springer, 2013.

Najafabadi, M. M.; Villanustre, F.; Khoshgoftaar, T. M.; Seliya, N; Wald, R.; Muharemagic, E. **Deep learning applications and challenges in *Big Data* analytics.** *Journal of Big Data.* Springer, 24 fev. 2015. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7>. Acesso em 25 out. 2024.

Nelson, R. **The market economy and the scientific commons.** *Research Policy* 33, p. 455-471, 2003. OCDE. Oslo Manual (Fourth Edition). Paris: OCDE, 2018.

Nextstrain. **Nextstrain: Real-time tracking of pathogen evolution.** 2024. Disponível em: <https://nextstrain.org>. Acesso em 30 out. 2024.

Noehrer, L.; Gilmore, A.; Jay, C.; Yehudi, Y. The impact of COVID-19 on digital data practices in museums and art galleries in the UK and the US. **Humanities and Social Sciences Communications**, v. 8(1), p. 1-10, 2021. <https://doi.org/10.1057/s41599-021-00921-8>.

OCDE. **EXPLANATORY MEMORANDUM ON THE UPDATED OECD DEFINITION OF AN AI SYSTEM.** OECD Artificial Intelligence Papers, March. 2024, no. 8. Disponível em: https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html. Acesso em 01 mai. 2025.

OCDE. **OECD Principles and Guidelines for Access to Research Data from Public Funding.** 2007. Disponível em: <https://web-archiver.oecd.org/2018-04-10/137520-38500813.pdf>. Acesso em 1 mai 2025.

Okediji, R. L. **Does Intellectual Property Need Human Rights?** 51 *NYU J. Int'L. Pol.* 1, 2018.

Okediji, R. L. **The Limits of Development Strategies at the Intersection of Intellectual Property and Human Rights.** In: Gervais, D. (ed.). **INTELLECTUAL PROPERTY, TRADE AND DEVELOPMENT.** 2ª ed., 2014, p. 355-361.

Okediji, R. L. **The Limits of International Copyright Exceptions for Developing Countries.** *Vanderbilt Journal of Entertainment and Technology Law*, 21, p. 689, 2020.

Disponível em: <https://scholarship.law.vanderbilt.edu/jetlaw/vol21/iss3/1>. Acesso em 09 jan. 2023.

Okediji, R. L. **The Regulation of Creativity Under the WIPO Internet Treaties**. Minnesota Legal Studies Research Paper, 09 (30), 2009.

Okediji, R. L. **WIPO-WTO Relations and the Future of Global Intellectual Property Norms**. Netherlands Yearbook of International Law, vol. 39, 2008.

Oliveira, J. G.; Rodrigues, F. C. R. DA TORRE DE MARFIM À UNIVERSIDADE EMPREENDEDORA: uma revisão de literatura sobre a Terceira missão acadêmica e o protagonismo no desenvolvimento socioeconômico. In: Amaral, M. G.; Mineiro, A. A. C.; Faria, A. F. (eds.). **AS HÉLICES DA INOVAÇÃO: interação universidade-empresa-governo-sociedade no Brasil**. Curitiba: CRV, 2022, p. 245-265.

Oliveira, J. V. **O FUTURO REPETINDO O PASSADO? DIGITAL RIGHTS MANAGEMENT, TECNOLOGIAS DISRUPTIVAS E O DIREITO AUTORAL BRASILEIRO**. Quaestio Iuris, vol. 12, (02), Rio de Janeiro, 2019, p. 647-672.

Oliveira, L. M. B. **Elementos estruturantes para a elaboração de políticas de direito autoral para repositórios institucionais de acesso aberto**. 2015. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília.

OMPI. Committee on Development and Intellectual Property (CDIP). **REVISED PILOT PROJECT ON TEXT AND DATA MINING (TDM) TO SUPPORT RESEARCH AND INNOVATION IN UNIVERSITIES AND OTHER RESEARCH-ORIENTED INSTITUTIONS IN AFRICA – PROPOSAL SUBMITTED BY THE AFRICAN GROUP (CDIP/30/9 REV)**. Genebra, 28 abr. 2023. 2023a. Disponível em: https://www.wipo.int/edocs/mdocs/mdocs/en/cdip_30/cdip_30_9_rev.pdf. Acesso em 28 set. 2024.

OMPI. **Convenção de Berna para a Proteção de Obras Artísticas e Literárias**. Paris, 24 jul. 1971. Última emenda em 28 set. 1979. Disponível em: <https://wipolex.wipo.int/en/text/283693>. Acesso em 25 out. 2024.

OMPI. **International Conference on Copyright Limitations and Exceptions for Libraries, Archives, Museums and Educational & Research Institutions**. Out. 2019. 2019a. Disponível em: https://www.wipo.int/meetings/en/2019/international_conference_copyright.html#archives. Acesso em 25 out. 2024.

OMPI. **Limitations and Exceptions**. 2024a. Disponível em: <https://www.wipo.int/copyright/en/limitations/>. Acesso em 29 out. 2024.

OMPI. **Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled.** 27 jun. 2013. Disponível em: <https://wipolex.wipo.int/en/text/301019>. Acesso em 25 out. 2024.

OMPI. **WIPO Copyright Treaty.** Genebra, 20 dez. 1996. Disponível em: http://www.WIPO.int/WIPOlex/en/treaties/text.jsp?file_id=295166. Acesso em: 09 out. 2019.

OMPI. **Generative AI. WIPO Conversation: IP and Frontier Technologies.** 2024b. Disponível em: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_conv_ge_2_23/wipo_ip_conv_ge_2_23_summary.pdf. Acesso em 20 set. 2024.

OMPI. Committee on Development and Intellectual Property (CDIP). **REVISED PILOT PROJECT ON TEXT AND DATA MINING (TDM) TO SUPPORT RESEARCH AND INNOVATION IN UNIVERSITIES AND OTHER RESEARCH-ORIENTED INSTITUTIONS IN AFRICA – PROPOSAL SUBMITTED BY THE AFRICAN GROUP (CDIP/30/9 REV.).** Genebra, 28 abr. 2023. 2023b. Disponível em: https://www.wipo.int/edocs/mdocs/mdocs/en/cdip_30/cdip_30_9_rev.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights. **WORK PROGRAM ON EXCEPTIONS AND LIMITATIONS (SCCR/43/8/REV.).** 17 mar. 2023. 2023c. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_43/sccr_43_8_rev.pdf. Acesso em 28 set. 2024.

OMPI. Standing Committee on Copyright and Related Rights. **DRAFT IMPLEMENTATION PLAN ON THE WORK PROGRAM ON LIMITATIONS AND EXCEPTIONS (SCCR/45/10 PROV.).** 19 abr. 2024. 2024c. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_45/sccr_45_10_prov.pdf. Acesso em 28 set. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Forty-Fifth Session. **INFORMATION SESSION ON THE OPPORTUNITIES AND CHALLENGES RAISED BY GENERATIVE AI - SCCR 45 DAY 4 AFTERNOON.** Genebra, 18 abr. 2024. 2024d. Disponível em: https://webcast.wipo.int/video/SCCR_45_2024-04-18_PM_122422. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Forty-Fourth Session. **DRAFT PROPOSAL BY THE AFRICAN GROUP FOR THE IMPLEMENTATION OF THE WORK PROGRAM ON EXCEPTIONS AND LIMITATIONS, ADOPTED AT THE 43RD SESSION OF THE WIPO SCCR, prepared by the African Group.** 3 nov. 2023. 2023d. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_44/sccr_44_6.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Forty-Fourth Session. **PROPOSAL FOR INFORMATION SESSION ON GENERATIVE AI AND COPYRIGHT prepared by Group B.** Genebra, 8 nov. 2023. 2023f. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_44/sccr_44_8.pdf. Acesso em 30 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Eighteenth Session. **Proposal by Brazil, Ecuador and Paraguay, Relating to Limitations and Exceptions: Treaty Proposed by the World Blind Union (WBU) (SCCR/18/5).** Genebra, 25 mai. 2009. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_18/sccr_18_5.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Twenty-Second Session. **Draft WIPO Treaty on Exceptions and Limitations for the Persons with Disabilities, Educational and Research Institutions, Libraries and Archives (SCCR/22/12).** Genebra, 3 jun. 2011. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_22/sccr_22_12.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Thirteenth Session. **Proposal by Chile on the Analysis of Exceptions and Limitations (SCCR/13/5).** Genebra, 22 nov. 2005. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_13/sccr_13_5.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Twenty-Fourth Session. **Limitations and Exceptions Regarding Education (SCCR/24/6).** Genebra, 16 jul. 2012. 2012a. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_24/sccr_24_6.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Twenty-Fourth Session. **Draft Articles and Thematic Clusters on Limitations and Exceptions to Copyright for the Benefit of Educational, Teaching and Research Institutions (SCCR/24/7).** Genebra, 16 jul. 2012. 2012b. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_24/sccr_24_7.pdf. Acesso em 29 out. 2024.

OMPI. Standing Committee on Copyright and Related Rights, Thirty-Sixth Session. **Action Plans on Limitations and Exceptions Through SCCR/39 (2nd Meeting in 2019) (SCCR/36/7).** Genebra, 1 jun. 2018. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_36/sccr_36_7.pdf. Acesso em 29 out. 2024.

OMPI. **The 45 Adopted Recommendations under the WIPO Development Agenda.** 2024e. Disponível em: <https://www.wipo.int/ip-development/en/agenda/recommendations.html#b>. Acesso em 29 out. 2024.

OMPI. **WIPO CONVERSATION ON INTELLECTUAL PROPERTY (IP) AND ARTIFICIAL INTELLIGENCE (AI) (WIPO/IP/AI/GE/19/1)**. 27 set. 2019. 2019b. Disponível em: https://www.wipo.int/edocs/mdocs/globalinfra/en/wipo_ip_ai_ge_19/wipo_ip_ai_ge_19_1.pdf. Acesso em 20 set. 2024.

OMPI. **WIPO CONVERSATION ON INTELLECTUAL PROPERTY (IP) AND ARTIFICIAL INTELLIGENCE (AI). Second Session**. 21 mai. 2020. Disponível em: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf. Acesso em 20 set. 2024.

OMPI. WIPO General Assembly, Thirty-First (15th Extraordinary) Session. **Proposal by Argentina and Brazil for the Establishment of a Development Agenda for WIPO (WO/GA/31/11)**. Genebra, 27 ago. 2004. Disponível em: https://www.wipo.int/edocs/mdocs/govbody/en/wo_ga_31/wo_ga_31_11.pdf. Acesso em 29 out. 2024.

OMPI. **WIPO Copyright Treaty**. 1996. Disponível em: https://www.wipo.int/wipolex/en/text/295166#P87_12240. Acesso em 28 set. 2024.

OpenAI. **About**. 2024a. Disponível em <https://openai.com/about>. Acesso em 28 fev. 2024.

OpenAI. **ChatGPT: Get answers. Find inspiration. Be more productive**. 2024b <https://openai.com/chatgpt/overview/>.

OpenAI. **DALL·E: Creating images from text**. 2024c. Disponível em <https://openai.com/research/dall-e>. Acesso em 28 fev. 2024.

OpenAI. **OpenAI developer platform**. 2024d. Disponível em: <https://platform.openai.com/docs/overview>. Acesso em 03 out. 2024.

OpenAI. **Our structure**. 2024e. Disponível em: <https://openai.com/our-structure/>. Acesso em 07 ago. 2024.

OpenAI. **Pioneering research on the path to AGI**. 2024f. Disponível em <https://openai.com/research/overview>. Acesso em 28 fev. 2024.

Open Government Data. **The Annotated 8 Principles of Open Government Data**. Disponível em: <https://opengovdata.org>. Acesso em 28 set. 2024.

Organização Mundial do Comércio. **Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (TRIPS)**. 1994. Disponível em: https://www.wto.org/english/docs_e/legal_e/27-trips.pdf. Acesso em 23 jun. 2022.

Organização Mundial Do Comércio. **WTO Analytical Index: TRIPS Agreement - Article 13 (Jurisprudence)**. Disponível em: https://www.wto.org/english/res_e/publications_e/ai17_e/trips_art13_jur.pdf. Acesso em 25 out. 2024.

Oriakhogba, D. O. **The Right to Research in Africa: Making African Copyright Whole**. Joint PIJIP/TLS Research Paper Series, ago. 2022.

Orsi, F.; Coriat, B. **The New Role and Status of Intellectual Property Rights in Contemporary Capitalism**. *Competition & Change*, v. 10 (2), pp. 162-179 (18), jun. 2006.

Ostrom, E. **Beyond Markets and States: Polycentric Governance of Complex Economic Systems**. *The American Economic Review*, vol. 100 (3), jun. 2010, p. 641-672.

Ostrom, E. **Reflections on the Commons**. Chapter 01. In: **Governing the commons: the evolution of institutions for collective action**. Cambridge University Press, 1990.

Ostrom, E.; Hess, C. **Private and Common Property Rights**. In: *Workshop in Political Theory and Policy Analysis*. Research Paper no. 2008-11-01. Indiana University, 2007.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K. Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P.; Leike, J.; Lowe, R. **Training language models to follow instructions with human feedback**. 04 mar. 2022. Disponível em <https://arxiv.org/abs/2203.02155>. Acesso em 28 fev. 2024.

Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D. et al. **PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews**. *BMJ*, 2021;372:n160, jan. 2021.

Pasquale, F. A.; Sun, H. **Consent and Compensation: Resolving Generative AI's Copyright Crisis**. Cornell Legal Studies Research Paper Forthcoming, University of Hong Kong Faculty of Law Research Paper no. 2024/07, 1 mai. 2024 (no prelo). Disponível em: <https://ssrn.com/abstract=4826695> ou <http://dx.doi.org/10.2139/ssrn.4826695>. Acesso em 04 out. 2024.

.

Perleman, H. S. **Taking the Protection-Access Tradeoff Seriously**. *VANDERBILT L REV.*, 53, p. 1831-1834, 2000.

Peters, J. HarperCollins is asking authors to license their books for AI training: Authors would have to opt-in to the agreement with an unnamed AI firm, with one reporting an offer of \$2,500 to license their book for three years. **The Verge**, 18 nov. 2024. Disponível em: <https://www.theverge.com/2024/11/18/24299882/harpercollins-authors-license-books-ai-training>. Acesso em 13 fev. 2025.

Pinheiro, A. M.; Tigre, p. B. (eds.). **Inovação em serviços na economia do compartilhamento**. Rio de Janeiro: Saraiva, 2019.

Pinheiro, P. P. G. **O direito internacional da propriedade intelectual aplicado à inteligência artificial**. 2018. Tese (Doutorado em Direito Internacional) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2018. doi:10.11606/T.2.2018.tde-23102020-005146.

Pinto, G. F. S. **COMPORTAMENTO INFORMACIONAL E MINERAÇÃO TEXTUAL NO TWITTER**. 2018. Dissertação (Ciência da Informação) – Faculdade de Ciência da Informação, Universidade Federal de São Carlos, São Carlos.

Rahmatian, A. **Neo-Colonial Aspects of Global Intellectual Property Protection**. *The Journal of World Intellectual Property*, vol. 12 (1), pp. 40–74, 2009.

Reino Unido. Intellectual Property Office. **Consultation outcome: Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation (28 jun. 2022)**. 2022a. Disponível em: <https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation>. Acesso em 31 ago. 2022.

Reino Unido. **Communications and Digital Committee. Corrected oral evidence: A creative future. Evidence Session No. 15. Heard in Public. Questions 131-142**. 22 nov. 2022. 2022b. Disponível em <https://committees.parliament.uk/oralevidence/11918/html/>. Acesso em 31 ago. 2024.

Reino Unido. **Oral Answers to Questions Volume 723**. 1 dez. 2022. 2022c. Disponível em <https://hansard.parliament.uk/commons/2022-12-01/debates/4AE6D1D0-7A3F-49A9-8634-AC7AED52BE15/OralAnswersToQuestions>. Acesso em 31 ago. 2024.

Reino Unido. **Copyright, Designs and Patents Act 1988 (Chapter 48, updated on October 12, 2023)**. Disponível em: <https://www.wipo.int/wipolex/en/legislation/details/22226>. Acesso em 28 set. 2024.

Renault, T. B.; Fonseca, M V. A.; Cunha, R. M.; Carvalho, R. S. Empreendedorismo acadêmico na COPPE/UFRJ: Reflexões sobre empresas criadas com a participação de professores. **Organizações em contexto**, v. 7 (14), p. 1-28, jul.-dez. 2011.

Rogers, S.; Schwabish, J.; Bowers, D. Data Journalism in 2017. **Google News Lab**, 2017. <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>.

Rosenberg, D. **Data before the Fact**. In: GITELMAN, L. **Raw Data is an Oxymoron**. The MIT Press, 2013.

Rowley, J. **The wisdom hierarchy: representations of the DIKW hierarchy**. *Journal of Information Science*, vol. 33 (2), pp. 163, 2007.

Russel, S.; Norvig, P. **Artificial Intelligence**. Traduzido por Regina Célia Simille 3rd ed. Elsevier, 2013.

Samuel, A. L. **Some Studies in Machine Learning Using the Game of Checkers**. *IBM Journal of Research and Development*, vol. 3 (3), pp. 210–229, jul. 1959.

Sag, M. **The New Legal Landscape for Text Mining and Machine Learning**. 66 *J. Copyright Soc’y USA* 1, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331606. Acesso em: 06 jan. 2023.

SAG-AFTRA. **Tentative Agreement Reached**. 8 nov. 2023. Disponível em: <https://web.archive.org/web/20231109204557/https://www.sagaftra.org/tentative-agreement-reached>. Acesso em 03 out. 2024.

Sagna, M. J. **STUDY ON THE DIGITAL MUSIC MARKET IN WEST AFRICA**. **OMPI, Standing Committee on Copyright and Related Rights Forty-First Session**. Genebra, 1 jun. 2021.

Samtani, S. Developing a Human Right to Research in International Law. **Joint PIJIP/TLS Research Paper Series**, 23 out. 2023.

Samuelson, P.; Asay, C. D. Saving Software’s Fair Use Future. **Harvard Journal of Law and Technology** 31, Special Issue Spring 2018, 1-28, 2018.

Santos, M. E. R.; Torkomian, A. L. V. Technology transfer and innovation: The role of the Brazilian TTOs. **International Journal of Technology Management & Sustainable Development**, v. 12, n. 1, p. 89-111, 2013. Disponível em: https://doi.org/10.1386/tmsd.12.1.89_1. Acesso em 27 out. 2024.

Savage, C. J., & Vickers, A. J. Empirical study of data sharing by authors publishing in PLoS journals. **PLoS one**, 4(9), e7078, 2009. Disponível em: <https://doi.org/10.1371/journal.pone.0007078>. Acesso em 28 set. 2024.

Schirru, L. A Inteligência Artificial e o Big Data no Setor da Saúde: Os Sistemas Especialistas e o Direito. **PIDCC**, Aracaju, vol. 10 (03), out. 2016, p. 81-99.

Schirru, L. **Direito Autoral e Inteligência Artificial: Autoria e titularidade nos produtos da IA**. 2020. Tese (Doutorado em Políticas Públicas, Estratégias e Desenvolvimento) – Instituto de Economia, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro.

Schirru, L., Souza, A. R., Valente, M. G.; Lana, A. P. Text and Data Mining Exceptions in Latin America. **IIC - International Review of Intellectual Property and Competition Law**, 19 set. 2024. Disponível em: <https://doi.org/10.1007/s40319-024-01511-2>. Acesso em 25 out. 2024.

Schirru, L.; Margoni, T. Arts 3 and 4 of the CDSM Directive as regulatory interfaces: Shaping contractual practices in the Commercial Scientific Publishing and Stock Images sectors. **Kluwer Copyright Blog**, 22 ago. 2023. Disponível em: <https://copyrightblog.kluweriplaw.com/2023/08/22/arts-3-and-4-of-the-cdsm-directive-as-regulatory-interfaces-shaping-contractual-practices-in-the-commercial-scientific-publishing-and-stock-images-sectors/>. Acesso em 24 set. 2024.

SciELO. **Crítérios, política e procedimentos para a admissão e a permanência de periódicos na Coleção SciELO Brasil**. Set. 2022. Disponível em: <https://www.scielo.br/media/files/20220900-criterios-scielo-Brasil.pdf>. Acesso em 28 set. 2024.

Scopus. **What is Scopus AI?** 15 ago. 2024. Disponível em: https://service.elsevier.com/app/answers/detail/a_id/36844/c/10543/supporthub/scopus/related/1/. Acesso em 27 set. 2024.

Sell, S. K. **Private Power, Public Law: The Globalization of Intellectual Property Rights**. Reino Unido: Cambridge University Press, 2003.

Senftleben, M. **AI Act and Author Remuneration - A Model for Other Regions?** 24 fev. 2024. Disponível em: <https://ssrn.com/abstract=4740268> ou <http://dx.doi.org/10.2139/ssrn.4740268>. Acesso em 04 out. 2024.

Senftleben, M. **Compliance of National TDM Rules with International Copyright Law – An Overrated Nonissue?** 2022. Disponível em: <https://ssrn.com/abstract=4134651> ou <http://dx.doi.org/10.2139/ssrn.4134651>. Acesso em: 6 jan. 2023.

Senffleben, M. **Copyright, limitations and the three -step test: an analysis of the three-step test in international and EC copyright law.** The Hague: Kluwer, 2004.

Senffleben, M. **Generative AI and Author Remuneration.** *International Review of Intellectual Property and Competition Law* 54, 2023, pp. 1535-1560. Disponível em: <https://ssrn.com/abstract=4478370> ou <http://dx.doi.org/10.2139/ssrn.4478370>. Acesso em 03 out. 2024.

Shaver, L. **The Right to Read.** *Columbia Journal of Transnational Law.* 54. 1-58, 2015a. Disponível em: <http://dx.doi.org/10.2139/ssrn.2467635>. Acesso em 28 out. 2024.

Shaver, L. **The right to science: ensuring that everyone benefits from scientific and technological progress.** Indiana: Indiana University Robert H. McKinney School of Law Research, 2015, Paper No. 2015-5. 2015b. Disponível em: <http://ssrn.com/abstract=2564222> or <http://dx.doi.org/10.2139/ssrn.2564222> Acesso em: 28 out. 2024.

Sindicato dos Trabalhadores na Indústria Cinematográfica e do Audiovisual dos Estados de SP, RS, MT, MS, GO, TO e DF (SINDCINE). **Ofício nº 20/2024-PRESIDÊNCIA. Assunto: Inclusão de temas envolvendo o setor Audiovisual Brasileiro no Projeto de Lei nº 2338, de 2023.** São Paulo, 8 abr. 2024. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9600139&ts=1720798354445&disposition=inline>. Acesso em 28 set. 2024.

Soares, M.N.; Kauffman, M.E.; Sales, G.M.C. **AVANÇOS DA COMUNIDADE EUROPÉIA NO DIREITO DE PROPRIEDADE INTELECTUAL E INDÚSTRIA 4.0: EXTRATERITORIALIDADE E APLICABILIDADE DO DIREITO COMPARADO NO BRASIL.** *Revista do Direito. Santa Cruz do Sul*, v. 1, n. 57, jan/abr. 2019, p. 117-137.

Sobel. B. L. W. Artificial Intelligence's Fair Use Crisis. 41 *Colum. J.L. & Arts* v. 45, 2017.

Sokol, D.; Comerford, R. E. **Antitrust and Regulating Big Data.** *UF Law Scholarship Repository*, 2016. Disponível em: <https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1817&context=facultypub>. Acesso em: 23 jun. 2022.

Solomon, D. J.; Björk, B.-C. A study of open access journals using article processing charges. *J Am Soc Inf Sci Tec*, 63, 2012, p. 1485-1495. Disponível em: <https://doi.org/10.1002/asi.22673>. Acesso em 23 out. 2024.

Somepalli G.; Singla, V.; Goldblum, M.; Geiping, J.; Goldstein, T. **Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models.** 12 dez. 2022. Disponível em: <http://arxiv.org/abs/2212.03860>. Acesso em 29 out. 2024.

Souza, A. R. **A função social dos direitos autorais: uma interpretação civil-constitucional dos limites da proteção jurídica. Brasil: 1988 – 2005.** 2005. Dissertação (Mestrado em Relações Privadas e Constituição) – Faculdade de Direito de Campos, Campos dos Goytacazes.

Souza, A. R.; Fairbanks, A. S. P. The Marrakesh Treaty Ratification In Brazil: Immediate Effects. **Panorama of Brazilian Law**, vol. 4, pp. 328-346, 2017.

Souza, A. R.; Schirru, L.; Alvarenga, M. B. **Direitos autorais e mineração de dados e textos no combate à COVID-19 no Brasil.** *LiinC em Revista*, vol. 16 (2), p. 1-15, 2020.

Souza, A. R.; Schirru, L.; Alvarenga, M. B. **Mineração de textos e dados na pesquisa em saúde: reflexões sobre direitos autorais.** *Cadernos de Saúde Pública*, v. 40 (5), 2024. DOI: <https://doi.org/10.1590/0102-311XPT169023>. Disponível em: <https://www.scielo.br/j/csp/a/b6RqfNLGGR6vdpbPJv5Gk6c/?lang=pt>. Acesso em 23 jan. 2025.

Souza, C. A. P. **Abuso do direito autoral.** 2009. Tese (Doutorado em Direito Civil) – Centro de Ciências Sociais, Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro.

Srnicek, N. **Platform capitalism.** Cambridge: Polity Press, 2017.

Stucke, M. E.; Grunes, A. P. **Debunking the Myths Over Big Data and Antitrust.** In: *CPI Antitrust Chronicle* (2) mai. 2015.

Quintais, J. P. **The New Copyright in the Digital Single Market Directive: A Critical Look.** *European Intellectual Property Review*, 42(1), 28-41, 2020. Disponível em: <https://doi.org/10.2139/ssrn.3424770>. Acesso em 22 jan. 2023.

Reichman, J. H.; Okediji, R. L. **When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale.** *Minn. L. Rev.* 1362, 1478, 2012. Disponível em: https://www.minnesotalawreview.org/wp-content/uploads/2019/07/ReichmanOkediji_MLR1362.pdf. Acesso em: 6 jan. 2023.

Singapura. **Copyright Act 2021 (Act No. 22 of 2021, amended by the Statutes (Miscellaneous Amendments) Act 2022).** Disponível em: <https://www.wipo.int/wipolex/en/legislation/details/21883>. Acesso em 29 out. 2024.

Suíça. **Federal Act of October 9, 1992, on Copyright and Related Rights (status as of January 1, 2022).** Disponível em: <https://www.wipo.int/wipolex/en/legislation/details/21313>. Acesso em 30 out. 2024.

Tennant, J. P.; Waldner, F.; Jacques, D. C.; Masuzzo, P.; Collister, L. B.; Hartgerink, C. H. J. **The academic, economic and societal impacts of Open Access: an evidence-based review.** *F1000Research*, 5:632, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837983/>. Acesso em 22 jun. 2022.

Tenopir, C.; Allard, S.; Douglass, K.; Aydinoglu, A. U.; Wu, L.; Read, E.; Manoff, M.; Frame, M. **Data Sharing by Scientists: Practices and Perceptions.** 2011. *PLoS One*, v. 6, n. 6, e21101. 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3126798/>. Acesso em 25 out. 2024.

Tenopir, C.; Dalton, E.; Allard, S.; Frame, M.; Piesivac, I.; Birch, B.; Pollock, D.; Dorsett, K. **Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide.** *PLoS One*, v. 10, n. 8, e0134826, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550246/>. Acesso em 25 out. 2024.

Tenopir, C.; Rice, N. M.; Allard, S.; Baird, L.; Borycz, J.; Christian, L.; Grant, B.; Olendorf, R.; Sandusky, R. J. **Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide.** 2020. *PLoS One* v. 15, n. 3, p. e0229003, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7065823/>. Acesso em 25 out. 2024.

The Next Rembrandt. **The Next Rembrandt.** 5 abr. 2016. Disponível em: <https://www.youtube.com/watch?v=IuygOYZ1Ngo>. Acesso em 30 out. 2024.

Torres, M; Xalabarder, R. **REPORT ON PRACTICES AND CHALLENGES IN RELATION TO ONLINE DISTANCE EDUCATION AND RESEARCH ACTIVITIES.** Organização Mundial da Propriedade Intelectual, Standing Committee on Copyright and Related Rights, Thirty-Ninth Session (SCCR/39/6). Genebra, 15 out. 2019. Disponível em: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_39/sccr_39_6.pdf. Acesso em 29 out. 2024.

Truyens, M.; van Eecke, P. **Legal aspects of text mining.** *Computer Law & Security Review*, vol. 30, 2014.

UBC. **Nos EUA, partidos rivais se unem em projeto para regular IA e proteger autores.** Madri, 17 jul. 2024. Disponível em <https://www.ubc.org.br/publicacoes/noticia/22655/nos-eua-partidos-rivais-se-unem-em-projeto-para-regular-ia-e-proteger-autores>. Acesso em 30 ago. 2024.

UC Berkeley. **What is Machine Learning (ML)?** 2020 (atualizado em fev. 2022). Disponível em: <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>. Acesso em 24 jul. 2024

Ueno, T. 'Rethinking the Provisions on Limitations of Rights in the Japanese Copyright Act: Toward the Japanese-style 'Fair Use' Clause'. 7 **Journal of the Japanese Group of AIPPI** 159, 2009.

União Europeia. Comissão Europeia. **Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões "Construir uma Economia Europeia dos Dados"**. Bruxelas, 10 jan. 2017. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52017DC0009&from=EN>. Acesso em 25 out. 2024.

União Europeia. Comissão Europeia. **Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões "Rumo a um Espaço Comum Europeu de Dados"**. Bruxelas, 25 abr. 2018. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52018DC0232&from=EN>. Acesso em 25 out. 2024.

União Europeia. **Diretiva 2001/29/CE do Parlamento Europeu e do Conselho, de 22 de Maio de 2001, relativa à harmonização de certos aspectos do direito de autor e dos direitos conexos na sociedade da informação**. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32001L0029>. Acesso em 29 out. 2024.

União Europeia. **Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho de 17 de abril de 2019, relativa aos direitos de autor e direitos conexos no Mercado Único Digital e que altera as Diretivas 96/9/CE e 2001/29/CE (Diretiva CDSM)**. 2019a. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32019L0790&from=PT>. Acesso em 25 out. 2024.

União Europeia. **Diretiva 96/9/CE do Parlamento Europeu e do Conselho, de 11 de março de 1996, relativa à proteção jurídica das bases de dados**. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:31996L0009&from=EN>. Acesso em 25 out. 2024.

União Europeia. **Mercado Único Digital: Suprimir os entraves para aproveitar as oportunidades em linha**. 2019b. Disponível em: https://ec.europa.eu/commission/priorities/digital-single-market_pt. Acesso em 07 nov. 2019.

União Europeia. PARLAMENTO EUROPEU, CONSELHO DA UNIÃO EUROPEIA. **DIRETIVA 2006/116/CE DO PARLAMENTO EUROPEU E DO CONSELHO de 12 de Dezembro de 2006, relativa ao prazo de protecção do direito de autor e de certos direitos conexos (versão codificada)**. Jornal da União Europeia, 27 dez. 2006. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32006L0116&from=EN>. Acesso em 22 abr. 2019.

União Europeia. REGULAMENTO (UE) 2024/1689 DO PARLAMENTO EUROPEU E DO CONSELHO de 13 de junho de 2024 que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento da Inteligência Artificial). Disponível em https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=OJ:L_202401689. Acesso em 30 ago. 2024.

Uruguai. Comisión de Educación y Cultura. **CARPETA N° 1923 DE 2017. DERECHOS DE AUTOR. Modificación de varias disposiciones de la Ley N° 9.739. Informe. ANEXO I AL REPARTIDO N° 50.** Mai. 2024. Disponível em: <https://documentos.diputados.gub.uy/docs/L49/Repartid/R0050-A01.pdf>. Acesso em 02 out. 2024.

Vyas, L **Taking Copyright's Balance Too Seriously**. *Indian J. Intell. Prop.* 13 (1), L 73, 2023.

Valente, M. G. **Reconstrução do Debate Legislativo sobre Direito Autoral no Brasil: Os anos 1989 – 1998**. 2018. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade de São Paulo, São Paulo.

Valente, M. G.; Alvarenga, M. B. **Mineração de dados no descompasso jurídico**. In: Santos, M. J. P.; Schaal, F. M. M.; Goulart, R. (orgs.). **A Propriedade Intelectual e o Mundo da Inteligência Artificial**. 1ed. PI & NewTech Editores, 2021.

Valente, M. G.; Lana, A. P.; Houang, A. P. **Copyright and Research in Latin America: law, courts, and perceptions**. *GRUR International*, 2024 (no prelo).

Van Dalen, H. P.; Henkens, K. **Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey**. *Journal of the American Society for Information Science and Technology*. 63, 2012, p. 1282-1293. Disponível em: [10.2139/ssrn.1983205](https://ssrn.com/abstract=1983205). Acesso em 28 out. 2024.

Wachowicz, M. A Revisão da Lei Brasileira de Direitos Autorais. In: Wachowicz, M.; Santos, M. J. P. (org). **Estudos de Direito de Autor: A Revisão da Lei de Direitos Autorais**. Florianópolis: Fundação Boiteux, 2010.

Widder, D.G.; Whittaker, M.; West, S.M. Why ‘open’ AI systems are actually closed, and why this matters. *Nature* 635, 2024, p. 827–833. Disponível em: <https://doi.org/10.1038/s41586-024-08141-1>. Acesso em 07 mar. 2025.

Wilson, L. **The Academic Man: A Study in the Sociology of a Profession**. Londres: Oxford University Press, 1942.

Woodmansee, M. The Genius and the Copyright: Economic and Legal Conditions of the Emergence of the 'Author'. **Eighteenth-Century Studies**, Vol. 17, No. 4, Special Issue: The Printed Word in the Eighteenth Century (Summer, 1984), p. 425-448.

Wouters, P.; Haak, W. **Open Data: The Researcher Perspective**. 2017. Disponível em: https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf. Acesso em 23 jun. 2022.

Wikimedia Enterprise. **Wikimedia Enterprise & Pleias Partner for Ethical AI Innovation**. 18 fev. 2025. Disponível em: <https://enterprise.wikimedia.com/blog/pleias-and-wikimedia-enterprise-partner/>. Acesso em 21 fev. 2025.

Wikimedia Foundation. **Wikimedia Enterprise/Essay**. Disponível em: https://meta.wikimedia.org/wiki/Wikimedia_Enterprise/Essay. Acesso em: 21 fev. 2025.

Wu, T. **Law and Economics of Information**. In: In: PARISI, F. (ed.). **The Oxford Handbook of Law and Economics: Volume 2: Private and Commercial Law**. Oxford Handbooks, 2017.

Xalabarder, R. **SCOPING STUDY ON THE PRACTICES AND CHALLENGES OF RESEARCH INSTITUTIONS AND RESEARCH PURPOSES IN RELATION TO COPYRIGHT**. Standing Committee on Copyright and Related Rights, Forty-Fourth Session. Geneva, 6-8 nov. 2023.

Yu, Peter K., Fair Use and Its Global Paradigm Evolution. **University of Illinois Law Review**, v. 2019, p. 111-69, **Texas A&M University School of Law Legal Studies Research Paper No. 18-24**, 02 jul. 2018. Disponível em: <https://ssrn.com/abstract=3206464>. Acesso em 30 out. 2024.

Zhou, C.; Etkowitz, H. Triple Helix Twins: A Framework for Achieving Innovation and UN Sustainable Development Goals. **Sustainability**, v. 13, n. 12, 2021. DOI: <https://doi.org/10.3390/su13126535>.

Zhuang, Y. **Studios and Writers Reached a Deal. Here's What Happens Next**. The New York Times, 25 set. 2023. Disponível em: <https://www.nytimes.com/2023/09/25/business/wga-writers-strike-deal-explained.html>. Acesso em 03 out. 2024.

APÊNDICE A – Exemplo de roteiro de entrevista

Instituição de pesquisa

1. Que materiais você usa nas suas investigações/pesquisas?
 - a. Você usa conteúdos/material de terceiros para pesquisas? De que tipo?
 - b. Qual a origem destes conteúdos/material?
 - c. Você sabe se elas são protegidas com Direito Autoral?
2. Você ou sua equipe fazem pesquisa com dados?
 - a. Como estes dados são obtidos?
 - b. Você sabe se essas bases de dados são de acesso aberto ou não?
 - c. Vocês precisam obter alguma autorização/licença para extrair e utilizar estes dados?
 - d. Você utiliza sistemas automatizados e/ou IAs para extrair e analisar os dados de pesquisa?
 - e. Você sabe se as bases de dados são protegidas por direitos autorais?
3. Como você entende o uso de conteúdos ou dados de terceiros?
 - a. Você acha que os dados de pesquisa são protegidos por alguma forma de propriedade (intelectual)? Deveriam ser? Qual? Quem seria o ‘dono’?
 - b. Você acha que é necessário autorização para utilizar conteúdos ou dados alheios para fins de pesquisa? Deveriam ser?
 - c. Quando você utiliza conteúdo ou dados alheios, Você se sente em risco?
 - d. O Direito Autoral (ou a propriedade intelectual) é ou já foi um problema para condução da pesquisa? E para divulgação de seus resultados?
 - e. Você acha que as bases de dados deveriam ser protegidas por direitos autorais (ou outro tipo de propriedade intelectual)? Como?
4. Vocês têm algum material informativo ou instrucional (guia, manual, etc.) para trabalhar com conteúdos ou dados alheios? Existe na sua instituição alguma política ou estratégia de uso de dados e conteúdo alheios? Você pode compartilhar?

APÊNDICE B – Quadro com as emendas sobre mineração de textos e dados ao Projeto de Lei nº 2.338/2023 até 10 de dezembro de 2024

Número da emenda	Proponente	Data	Conteúdo	Aprovada?
24	Sen. Laércio Oliveira (PP/SE)	11/06/2024	Suprime o art. 59 e o art. 60, II	Retirada pelo autor em 14/06/2024
44	Sen. Chico Rodrigues (PSB/RR)	17/06/2024	Altera o art. 59 para obrigar desenvolvedores ao invés de aplicadores de IA, visando retirar a responsabilidade legal sobre os usuários	Sim
45	Sen. Chico Rodrigues (PSB/RR)	17/06/2024	Altera o art. 63, que dispõe sobre sandbox regulatório, com a competência direcionada para o SAI e a remoção da ressalva a usos não-comerciais	Parcialmente
55	Sen. Fabiano Contarato (PT/ES)	18/06/2024	Acrescenta o art. 64-A, que excluiria da limitação no art. 60 as instituições cobertas que passassem a explorar uma IA treinada com obras protegidas para fins lucrativos, com a responsabilidade legal pelo uso não-autorizado retroagindo até o período de uso não-lucrativo previamente protegido pela limitação ²⁵⁷	Não

²⁵⁷ “Acrescente-se à Seção IV do Capítulo X desta Lei (Direitos Autorais) o seguinte dispositivo:

Art. 64-A. As instituições de que trata o caput do art. 60 que passem a explorar sistemas de inteligência artificial com finalidade lucrativa, ou que passem a incidir na hipótese prevista no §2º do art. 60, sujeitam-se a todos os dispositivos desta Lei, em particular aos Arts. X, 59 e 63, inclusive quanto ao período anterior considerado sem fins lucrativos em que os conteúdos de direitos de autor e direitos conexos tenham sido utilizados para treinar e desenvolver os sistemas de inteligência artificial.

JUSTIFICAÇÃO

56	Sen. Weverton (PDT/MA)	18/06/2024	Modifica o art. 63 para adicionar uma cláusula de arbitragem	Não
61	Sen. Astronauta Marcos Pontes (PL/SP)	18/06/2024	Remove toda a seção sobre direitos autorais	Não
72	Sen. Astronauta Marcos Pontes (PL/SP)	18/06/2024	Propõe um novo texto para a seção de Direito Autoral, com destaque para o art. 60.	Não
109	Sen. Izalci Lucas (PL/DF)	02/07/2024	Adiciona cláusulas de negociação contratual e reciprocidade internacional ao art. 64 ²⁵⁸²⁵⁹	Não

A proposta busca abranger empresas que nascem sem fins lucrativos, treinam seus modelos, desenvolvem e posteriormente mudam seu regime para fins lucrativos. Os casos podem ser comuns, tendo em vista que muitas startups nascem em ambientes de pesquisa ou que a priori não visam apresentar-se ao mercado. Não podem, todavia, caso decidam depois de produzido o produto, alterar sua proposta para auferir lucros, deixar de fora a proteção dos direitos autorais.”

²⁵⁸ “Art. 64.

(...)

Parágrafo único

(...)

III - a livre negociação na utilização dos conteúdos protegidos, com promoção de um ambiente de pesquisa e experimentação que possibilite o desenvolvimento de práticas inovadoras, e que não restrinjam a liberdade de pactuação entre as partes envolvidas.

(...)

V – (...)

b) a pessoas domiciliadas em país que assegure a reciprocidade na proteção, em termos equivalentes a este artigo, aos direitos de autor e aos direitos conexos de brasileiros, conforme disposto no artigo 2º, parágrafo único, da Lei n. 9.610, de 1998, sendo vedada a cobrança nos casos em que a reciprocidade não estiver assegurada.

JUSTIFICAÇÃO

No que se refere ao inciso III, propõe-se remover as menções ao Código Civil e à Lei de Direitos Autorais. Quanto à Lei de Direitos Autorais, a menção é desnecessária, visto que a temática de direitos autorais já é regida pela referida legislação, que continuaria sendo aplicável aos casos envolvendo inteligência artificial e direitos autorais. Quanto ao Código Civil, os referidos dispositivos buscam inserir, no contexto dos contratos envolvendo inteligência artificial e direitos autorais, normas que permitam a resolução ou revisão contratual.

(...)

Por fim, o ajuste na alínea b do inciso V do artigo faz-se necessário para evitar que a regra de reciprocidade esteja condicionada à existência de sociedades de gestão coletiva no país estrangeiro, o que impediria que titulares estrangeiros pudessem exercer seus direitos individualmente mesmo quando a reciprocidade é assegurada.”

²⁵⁹ Aqui, cabe menção às Emendas 110, do Sen. Carlos Portinho (PL/RJ), 118, do Sen. Mecias de Jesus (REPUBLICANOS/RR) e 128, do Sen. Astronauta Marcos Pontes (PL/SP), que previam texto similar e foram igualmente rejeitadas pelo Senado, razão pela qual não foram incluídas no quadro.

109	Sen. Izalci Lucas (PL/DF)	02/07/2024	Adiciona cláusulas de negociação contratual e reciprocidade internacional ao art. 64 ²⁶⁰²⁶¹	Não
151	Sen. Marcos Rogério (PL/RO)	29/11/2024	Remove toda a seção sobre direitos autorais ²⁶²	Parcialmente
159	Sen. Carlos Portinho (PL/RJ)	02/12/2024	Modifica os arts. 62 a 64 do dispositivo suprime o art. sobre uso de imagem	Parcialmente
161	Sen. Mecias de Jesus (REPUBLICANOS/RR)	02/12/2024	Modifica os arts. 62 a 64 do dispositivo, suprimindo o art. sobre uso de imagem	Não

²⁶⁰ “Art. 64.

(...)

Parágrafo único

(...)

III - a livre negociação na utilização dos conteúdos protegidos, com promoção de um ambiente de pesquisa e experimentação que possibilite o desenvolvimento de práticas inovadoras, e que não restrinjam a liberdade de pactuação entre as partes envolvidas.

(...)

V – (...)

b) a pessoas domiciliadas em país que assegure a reciprocidade na proteção, em termos equivalentes a este artigo, aos direitos de autor e aos direitos conexos de brasileiros, conforme disposto no artigo 2º, parágrafo único, da Lei n. 9.610, de 1998, sendo vedada a cobrança nos casos em que a reciprocidade não estiver assegurada.

JUSTIFICAÇÃO

No que se refere ao inciso III, propõe-se remover as menções ao Código Civil e à Lei de Direitos Autorais. Quanto à Lei de Direitos Autorais, a menção é desnecessária, visto que a temática de direitos autorais já é regida pela referida legislação, que continuaria sendo aplicável aos casos envolvendo inteligência artificial e direitos autorais. Quanto ao Código Civil, os referidos dispositivos buscam inserir, no contexto dos contratos envolvendo inteligência artificial e direitos autorais, normas que permitam a resolução ou revisão contratual.

(...)

Por fim, o ajuste na alínea b do inciso V do artigo faz-se necessário para evitar que a regra de reciprocidade esteja condicionada à existência de sociedades de gestão coletiva no país estrangeiro, o que impediria que titulares estrangeiros pudessem exercer seus direitos individualmente mesmo quando a reciprocidade é assegurada.”

²⁶¹ Aqui, cabe menção às Emendas 110, do Sen. Carlos Portinho (PL/RJ), 118, do Sen. Mecias de Jesus (REPUBLICANOS/RR) e 128, do Sen. Astronauta Marcos Pontes (PL/SP), que previam texto similar e foram igualmente rejeitadas pelo Senado, razão pela qual não foram incluídas no quadro.

²⁶² “JUSTIFICAÇÃO

A retirada da discussão sobre direito autoral se justifica, primeiramente, pela complexidade do tema, que entendemos demandar uma análise detalhada e específica, com nuances técnicas e legais que não podem ser adequadamente abordadas em um debate generalista. É um assunto que exige um espaço próprio, em que todas as implicações possam ser exploradas de forma mais profunda e cuidadosa.

(...)

Sobretudo, devemos considerar o impacto que uma regulamentação restritiva poderia ter no cotidiano dos brasileiros, especialmente no uso de ferramentas digitais que já são amplamente adotadas. Muitas dessas tecnologias, que auxiliam tanto no trabalho quanto no lazer, poderiam ser comprometidas por uma aplicação rigorosa do direito autoral, criando obstáculos ao acesso, à inclusão digital, à acessibilidade de portadores de necessidades específicas e à inovação.” (Brasil, 2024c).

162	Sen. Mecias de Jesus (REPUBLICANOS/RR)	02/12/2024	Modifica os arts. 62 e 63 do dispositivo e suprime os arts. sobre <i>sandbox</i> regulatório e uso de imagem	Parcialmente
168	Sen. Mecias de Jesus (REPUBLICANOS/RR)	02/12/2024	Modifica os arts. 63 a 65 do dispositivo ²⁶³	Não
178	Sen. Mecias de Jesus (REPUBLICANOS/RR)	02/12/2024	Modifica o art. sobre remuneração restringindo-a a IAs com fins comerciais e estabelecendo data limite para sua aplicação ²⁶⁴	Não
185	Sen. Izalci Lucas (PL/DF)	02/12/2024	Modifica os arts. Sobre transparência e mineração de textos e dados no dispositivo e suprime arts. sobre remuneração e uso de imagem	Parcialmente
187	Sen. Izalci Lucas (PL/DF)	02/12/2024	Modifica o art. sobre transparência do substitutivo e suprime o art. sobre uso de imagem	Parcialmente
191	Sen. Efraim Filho (UNIÃO/PB)	02/12/2024	Modifica a cláusula de transparência do substitutivo, estabelecendo meios para a informação de todos os conteúdos protegidos usados em IA	Sim

²⁶³ Destaque para a disposição de reserva legal a seguir: “Art. 64. O autor ou artista, quando da disponibilização da obra em bancos de dados públicos ou na rede mundial de computadores, tem a faculdade de utilizar ferramentas tecnológicas que impeçam a mineração de texto e dados no desenvolvimento de sistemas de IA com fins comerciais, na forma do regulamento setorial.

§ 1o A faculdade a que se refere o caput desse artigo poderá ser exercida pelo autor, artista ou criador da obra, pelo uso das tecnologias disponíveis e indicadas pelo órgão regulador, conforme o estado da arte.

§ 2o O exercício da faculdade prevista no caput, não retroage para alcançar o momento anterior à disponibilização da obra com a utilização da ferramenta tecnológica prevista no caput e no § 1o, desse artigo.” (Brasil, 2024c).

²⁶⁴ “JUSTIFICAÇÃO (...) Em relação ao parágrafo único, é importante esclarecer que, tecnicamente, um sistema de IA não tem como ser destreinado. Assim, faz-se necessário deixar inequívoco no projeto de lei que a proibição de uso de obras protegidas surte efeitos a partir do momento em que for expressa.” (Brasil, 2024c).

192	Sen. Efraim Filho (UNIÃO/PB)	02/12/2024	Modifica a limitação para mineração de textos e dados, deixando mais clara a permissão para processos de mineração e impondo proibições a fins comerciais, jornalísticos e à disseminação das obras utilizadas	Parcialmente
193	Sen. Efraim Filho (UNIÃO/PB)	02/12/2024	Modifica a reserva legal para uso em IA para permitir sua implementação por qualquer meio, inclusive simples notificação	Não
194	Sen. Efraim Filho (UNIÃO/PB)	02/12/2024	Suprime o artigo sobre remuneração	Parcialmente
200	Sen. Esperidião Amin (PP/SC)	05/12/2024	Introduz novo texto para as limitações e proibições para treinamento de IA ²⁶⁵	Não
220	Sen. Mecias de Jesus (REPUBLICANOS/RR)	05/12/2024	Modifica a reserva legal para uso em IA para permitir sua implementação facilmente, de forma individualizada, por meios eletrônicos ²⁶⁶	Não

²⁶⁵ “**Art. 62.** O agente de IA poderá utilizar obras protegidas por direitos de autor e conexos no seu desenvolvimento, sujeito às seguintes condições:

I – a utilização da obra não tenha como objetivo principal a reprodução, exibição ou disseminação da própria obra original;

II – o agente de IA não compartilhe uma cópia da obra com quaisquer terceiros, exceto para fins de verificação dos resultados da análise de dados computacionais realizada pelo desenvolvedor, ou de pesquisa colaborativa ou estudos relacionados às finalidades da análise de dados computacionais realizada pelo desenvolvedor;

III – o agente de IA tenha legalmente acessado o material usado no desenvolvimento.”

“**Art. 63.** Não constitui ofensa aos direitos de autor e conexos a utilização automatizada de conteúdos protegidos em processos de mineração de textos e de dados para os fins de pesquisa ou desenvolvimento de sistemas de IA, desde que eventuais cópias de conteúdos protegidos por direitos de autor e conexos utilizadas no treinamento de sistemas de IA sejam mantidas em estritas condições de segurança e unicamente pelo tempo necessário para a realização da atividade ou para a finalidade específica de verificação dos resultados.

Parágrafo único. A atividade de mineração de textos e de dados que envolva dados pessoais estará sujeita às disposições da Lei no 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais).”

“**Art. 64.** O titular de direitos de autor e conexos poderá optar por reservar seus direitos sobre as obras para impedir a mineração de textos e de dados para o desenvolvimento de sistemas de IA, exceto quando esta for feita para fins de pesquisa científica e desde que esta reserva seja realizada por meio de métodos compreensíveis por sistemas computacionais.

Parágrafo único. Nos casos em que a reserva de direitos tiver sido realizada expressa e validamente de forma adequada, e quando nenhuma outra exceção de uso seja aplicável, o agente de IA precisará obter autorização dos titulares de direitos autorais e conexos para a realização de mineração de texto e dados sobre as obras.”

²⁶⁶ “Art. 64

(...)

221	Sen. Mecias de Jesus (REPUBLICANOS/RR)	05/12/2024	Modifica os critérios para cálculo da remuneração devida pelo treinamento de IA ²⁶⁷	Não
225	Sen. Esperidião Amin (PP/SC)	05/12/2024	Modifica a redação na limitação para mineração de textos e dados para combate a ilícitos contra DA, visando proibir interpretações mais amplas ²⁶⁸	Não
229	Sen. Carlos Portinho (PL/RJ)	09/12/2024	Similar à Emenda nº 200	Não

§2º Os titulares devem poder exercer facilmente o direito previsto no *caput* deste artigo, para cada obra individualmente e por meios automatizados.”

²⁶⁷ “Art. 65 (...)

§ 1º (...)

II - que o cálculo da remuneração a que se refere o *caput*, na hipótese de gestão coletiva, considere os princípios da razoabilidade e da proporcionalidade e elementos relevantes, tais como a complexidade do sistema de IA desenvolvido, o porte do agente de IA, o ciclo de realização econômica dos sistemas de IA, o grau de utilização dos conteúdos, o valor relativo da obra ao longo do tempo e os efeitos concorrenciais dos resultados em relação aos conteúdos originais utilizados;

(...)

JUSTIFICAÇÃO

A redação mais recente do referido inciso introduz critérios destinados a orientar a formação de preços e condições de remuneração pelo uso de obras protegidas no desenvolvimento e treinamento de sistemas de IA.

Embora tais critérios sejam prática comum e amplamente aceita no âmbito da gestão coletiva de direitos autorais, eles não encontram paralelo nas negociações diretas e individuais de direitos – cuja possibilidade está prevista expressamente no inciso II do mesmo parágrafo.

Isso se justifica pelo fato de que a aplicação de critérios predeterminados a negociações individuais representaria uma limitação à autonomia das partes envolvidas e à liberdade de pactuação, princípios fundamentais no regime de direitos autorais.”

²⁶⁸ Em oposição, a exceção criada pelo parágrafo 4º do art. 63 é excessivamente ampla (“combate a ilícitos, civis e criminais, que atentem contra direitos de autor e conexos”) e não atende aos critérios da Convenção de Berna, bem como às demais restrições originalmente previstas pelo artigo. Ademais, o artigo não estabelece restrições quanto ao uso dos dados extraídos nesse contexto. Assim, mesmo que a coleta de dados ocorra em um “contexto de sistemas de inteligência artificial para combate a ilícitos”, a utilização desses dados não é limitada pela legislação proposta e pode ser destinada a qualquer finalidade comercial.”