

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO**

**APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL AO DIREITO: ANÁLISE DE
PARÂMETROS DA LITERATURA, DA REGULAÇÃO E DA PERCEPÇÃO DE
ATORES SOBRE LIMITES ÉTICOS.**

Brasília
2022

TAINÁ AGUIAR JUNQUILHO

**APLICAÇÃO DE INTELIGÊNCIA ARTIFICIAL AO DIREITO: ANÁLISE DE
PARÂMETROS DA LITERATURA, DA REGULAÇÃO E DA PERCEPÇÃO DE
ATORES SOBRE LIMITES ÉTICOS.**

Projeto de Tese de Doutorado apresentado ao Programa de Pós-Graduação em Direito da Universidade de Brasília como requisito para a obtenção do título de Doutora.

Linha de pesquisa: Constituição e Democracia
Orientadora: Prof.^a Dra. Cláudia Rosane Roesler
Coorientadora: Prof.^a Dra. Ana Cláudia Farranha

Brasília
2022

AGRADECIMENTOS

Preciso agradecer a algumas pessoas que foram fundamentais nessa caminhada até aqui.

À Força estranha cantada por Caetano, presente no ar que proporciona novos encontros, alegrias e lugares.

Aos meus pais, que me apoiaram desde o dia em que decidi me mudar para o cerrado, lugar pelo qual me apaixonei.

À minha avó e às famílias Aguiar e Junquilha: amores incondicionais.

À minha orientadora, Cláudia Rosane Roesler. Mulher forte que desde o primeiro momento que vi na UnB passei a admirar.

À Vera Roesler, suporte para sanidade do psicofísico.

À Eneá, que plantou a sementinha da vinda à Brasília e à família candanga que tanto amo: Bittencourt e Ninian.

A Elias Canal Freitas, meu grande amor.

Às amigas capixabas, em especial, Rosana, Mary, Lu Pianca, Helana, Rovená e Thayla, que me aguentaram à distância.

À Mariana Kuster e Luma Polleti, que me apoiaram tanto nas angústias e nas alegrias em Brasília.

À Mila Fregona, Júlia Fregona, Antônio Salgado, Jéssica Brágio, Rosa Burtet: minhas pérolas feitas pela ostra feliz.

Aos colegas do Projeto Victor, que tanto me ensinaram, em nome dos amigos Mamede e Nilton.

Aos colegas de Neoway, em especial, Bruna Müller, Ricardo Fernandes e Guilherme Sardinha, que me apoiaram tanto e acreditaram em mim.

À Petruska e família, que me apresentou Mira e sua família que tão bem me acolheu em terras secas.

A todos os entrevistados, que dedicaram um tempo dos seus dias corridos para o engrandecimento dessa pesquisa com contribuições valiosas.

Aos amigos que fiz no GPRAJ, que me apoiaram em tantas discussões e desabafos acadêmicos, em especial à Clélia e Isaac Reis.

Pela alegria da alma, resistência e inspiração para seguir: agradeço a minha Coorientadora, Ana Cláudia Farranha e à Teresa Cristina e suas preciosas *lives*.

A toda equipe da secretaria de pós em Direito da UnB, agradecimento que faço em nome de Euzilene.

Aos profissionais da saúde meu agradecimento por sua luta diária.

“Chegar para agradecer e louvar.

Louvar o ventre que me gerou

O orixá que me tomou,

E a mão da doçura de Oxum que consagrou.

Louvar a água de minha terra

O chão que me sustenta, o palco, o massapê,

A beira do abismo,

O punhal do susto de cada dia.

Agradecer as nuvens que logo são chuva,

Sereniza os sentidos

E ensina a vida a reviver.

Agradecer os amigos que fiz

E que mantém a coragem de gostar de mim, apesar de mim...
Agradecer a alegria das crianças,
As borboletas que brincam em meus quintais, reais ou não.
Agradecer a cada folha, a toda raiz, as pedras majestosas
E as pequeninas como eu, em Aruanda.
Agradecer o sol que raia o dia,
A lua que como o menino Deus espraia luz
E vira os meus sonhos de pernas pro ar.
Agradecer as marés altas
E também aquelas que levam para outros costados todos os males.
Agradecer a tudo que canta no ar,
Dentro do mato sobre o mar,
As vozes que soam de cordas tênues e partem cristais.
Agradecer os senhores que acolhem e aplaudem esse milagre.
Agradecer,
Ter o que agradecer. Louvar e abraçar! ”

Maria Bethânia.

Para meus pais.
E quem mais esteve comigo nessa caminhada (que por pura sorte minha não foram poucos).

RESUMO

A Inteligência Artificial, conceito que envolve variados modelos tecnológicos de desenvolvimento, vive seu *boom* em todas as áreas do conhecimento humano. A despeito de sua aplicação gerar inúmeros benefícios e avanços, pode trazer como consequências algumas externalidades negativas como a opacidade, os vieses discriminatórios e a violação aos dados pessoais e ao direito à privacidade. Por isso é direito fundamental exigir sua aplicação ética, isso é, com limitações que controlem essas externalidades. Essa pesquisa buscou compreender quais parâmetros éticos podem ser identificados e considerados no desenvolvimento da IA com foco na área do Direito. Por meio da triangulação dos resultados obtidos da literatura e apontados nas diretrizes atuais específicas para IA aplicada ao Direito, bem como pelos atores especialistas no tema que foram entrevistados, concluiu-se que o mínimo ético existente são os princípios que servem para aplicação de IA em todas as áreas: “Transparência”, “Responsabilidade”, “Equidade”, “Prestação de contas”, “Explicabilidade” e “Sustentabilidade” (acrônimo em português TREPES). Embora se reconheça a importância do consenso nos princípios, recomenda-se a criação de regulação vinculante geral, que atribua responsabilidades e regramentos setoriais e que classifique as aplicações da IA em riscos de acordo com suas especificidades.

PALAVRAS-CHAVE: IA no Direito; IA na Justiça; aprendizado de máquina; *design* ético da IA; princípios; diretrizes éticas

ABSTRACT

The concept of Artificial Intelligence involves several technological development models and is experiencing its boom in all areas of human knowledge. Despite its application generates countless benefits and advances, it can bring as consequences some negative externalities such as opacity, discriminatory biases and the violation of personal data and to the right to privacy. Therefore, it is a fundamental right to demand its ethical application. This research sought to understand which ethical parameters can be identified and considered in the development of AI focusing on the area of Law. Through the triangulation of the results obtained from the literature and pointed out in the current specific guidelines for AI applied to Law, as well by the experts on the subject who were interviewed, it was concluded that the existing ethical minimum are the principles that serve to apply AI in all areas: “Transparency”, “Responsibility”, “Equity”, “Accountability”, “Explicability” and “Sustainability” (acronym in Portuguese TREPES). Although the importance of consensus on the principles is recognized, the creation of general binding regulation is recommended, which assigns responsibilities and sectorial rules that classify AI applications in risks according to their specificities.

KEYWORDS: AI to Law; AI to Justice; Machine Learning; Ethical Design of AI; Principles; Ethical Guidelines.

LISTA DE FIGURAS

Figura 1- Linha do tempo IA.....	21
Figura 2- Geração automática de legendas de texto para imagens com redes profundas	26
Figura 3- Abordagens de Aprendizado de Máquina e técnicas.....	27
Figura 4- Aprendizado de Máquina e utilizações por técnica.....	29
Figura 5- Processo de Pesquisa em Inovação Responsável.....	40
Figura 6- Nuvem de palavras mais recorrentes nos textos completos a partir da pesquisa com descritores “ethics” and “artificial intelligence”.....	41
Figura 7- Gráfico de palavras recorrentes nos textos completos a partir dos descritores “ethics” and “artificial intelligence”.....	42
Figura 8- Nuvem de palavras-chave mais recorrentes a partir da pesquisa com descritores “ethics” and “artificial intelligence”.....	44
Figura 9- Gráfico de palavras-chave recorrentes a partir dos descritores “ethics” and “artificial intelligence”.....	44
Figura 10- Nuvem de palavras mais recorrentes a partir dos textos completos da pesquisa com descritores “artificial intelligence”, “decision making” and “law”.....	46
Figura 11- Gráfico de palavras recorrentes nos textos completos a partir dos descritores “artificial intelligence”, “decision making” and “law”.....	46
Figura 12- Nuvem de palavras-chave mais recorrentes a partir da pesquisa com descritores “artificial intelligence”, “decision making” and “law”.....	48
Figura 13- Gráfico de palavras-chave recorrentes a partir dos descritores “artificial intelligence”, “decision making” and “law”.....	48
Figura 14 - Nuvem de palavras mais recorrentes a partir dos textos completos da pesquisa com descritores “artificial intelligence” and “principle” and “Law”.....	49
Figura 15- Gráfico de palavras recorrentes nos textos completos a partir dos descritores “artificial intelligence” and “principle” and “Law”.....	50
Figura 16- Nuvem de palavras-chave mais recorrentes a partir da pesquisa com descritores “artificial intelligence” and “principle” and “Law”.....	52
Figura 17- Gráfico de palavra-chaves recorrentes a partir dos descritores “artificial intelligence” and “principle” and “Law”.....	52
Figura 18- Exemplo de funcionamento de algoritmos de aprendizagem profunda.....	122
Figura 19- Reformas de decisões nas câmaras criminais do TJSP.....	128
Figura 20- Representação da triangulação dos resultados.....	137

LISTA DE QUADROS

Quadro 1– Lista de usos de aprendizado de máquina no Direito	71
Quadro 2- Principais desafios colocados pela aplicação de IA ao Direito	82
Quadro 3– Análise das aplicações de IA no Poder Judiciário Brasileiro	84
Quadro 4- Lista não exaustiva dos serviços jurídicos que recorrem à inteligência artificial na área jurídica.	97
Quadro 5– Perfil dos entrevistados.....	106
Quadro 6– Categorias, subcategorias e autores de referência.	110
Quadro 7 – Resultado da triangulação.....	140

LISTA DE ABREVIATURAS

Aprendizado de Máquina (AM)

Associação Brasileira de Jurimetria (ABJ)

Associação Brasileira de *Lawtechs* e *Legaltechs* (AB2L)

Best Alternative To a Negotiated Agreement (BATNA)

Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais (CEEUIA SJ)

CEPEJ (Comissão Europeia para Eficiência da Justiça)

COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*)

Conselho Nacional de Justiça (CNJ)

Código de Processo Civil de 2015 (CPC/15)

Estratégia Brasileira de Inteligência Artificial (EBIA)

GDPR (General Data Protection Regulation)

Cadastro de Pessoas Físicas (CPF)

Generative Pre-Training Transformer (GPT)

Estratégia Brasileira de Inteligência Artificial (EBIA)

Fundação Getulio Vargas (FGV)

International Business Machines Corporation (IBM)

Lei Geral de Proteção de Dados (LGPD)

Laboratório de Inovação do Tribunal Federal da 3ª Região (ILabTRF3)

Laboratório de Inteligência Artificial da Universidade de Brasília (AI.LAb)

Laboratório de Inovação e Inteligência e Objetivos de Desenvolvimento Sustentável (LIODS)

Multilayer Perceptron (MLP)

Organização das Nações Unidas (ONU)

Optical Character Reader (OCR)

Organização Não Governamental (ONG)

Pesquisa Inovadora Responsável (PRI ou em inglês, RRI – *Responsible Research and Innovation*)

Plataforma Digital do Poder Judiciário Brasileiro (PDPJ-Br)

Processamento de Linguagem Natural (PLN ou Natural Language Process - NLP)

Processo Judicial Eletrônico (PJE)

Quantitative Legal Prediction (QLP)

Predição legal quantitativa (PLQ)

Redes Neurais Recorrentes (RNN)

Redes Neurais Convolucionais (CNN)

Web of Science (WoS)

Resolução Online de Disputas (*ODR – Online Disupute Resolution*)

Tecnologia de Propósito Geral (TGP)

Comissão Europeia para a Eficácia da Justiça (CEPEJ)

Tribunal de Justiça de São Paulo (TJSP)

União Europeia (UE)

SUMÁRIO

INTRODUÇÃO.....	14
1 INTELIGÊNCIA ARTIFICIAL: CONFUSÃO DO TERMO E EXIGÊNCIAS ÉTICAS	20
1.1.1 Compreendendo a história da IA	20
1.1.2 Pressuposto: o que é IA para essa pesquisa?	23
1.2 POR QUE É (DIREITO) FUNDAMENTAL EXIGIR APLICAÇÃO ÉTICA DA IA?	30
2 ÉTICA NA IA: OS <i>INPUTS</i> DA LITERATURA.....	40
3 <i>INPUTS</i> ÉTICOS NA INTELIGÊNCIA ARTIFICIAL: COMPARANDO DIRETRIZES	55
3.1 GOVERNANÇA DA IA APLICADA AO DIREITO	62
3.1.1 Usos da IA no Direito: concretização dos mandamentos do CPC/15	63
3.1.2. Questões éticas	72
3.1.3 IA NA JUSTIÇA BRASILEIRA.....	83
3.1.4 O que trazem as diretrizes Europeia e Brasileira?	94
4 <i>INPUTS</i> ÉTICOS APONTADOS PELOS GRUPOS DE ATORES ENVOLVIDOS COM INTELIGÊNCIA ARTIFICIAL.....	104
4.1 METODOLOGIA DE INVESTIGAÇÃO PARA AVERIGUAÇÃO DA PERCEPÇÃO DOS ATORES SOCIAIS	104
4.2 SELEÇÃO DOS ENTREVISTADOS.....	105
4.3 ASPECTOS ÉTICOS DA PESQUISA	108
4.4 ROTEIRO E APLICAÇÃO DAS ENTREVISTAS.....	108
4.5 ANÁLISE DOS DADOS: “O MEU PAÍS É O MEU LUGAR DE FALA”	109
4.5.1 IA no Direito.....	111
4.5.2 Ética na IA como pretensões principiológicas	115
4.5.3 “Desses que vivem no escuro em plena luz”	117
4.5.4 Dadaísmo judicial	123
4.5.5 Vieses como “limonada pronta”	127
4.5.6 “E agora, José?”	131
5 OUTPUT: TRIANGULAÇÃO DOS RESULTADOS.....	136
CONCLUSÕES	143
REFERÊNCIAS	147
APÊNDICES	157
Termo de Consentimento Livre e Esclarecido (TCLE).....	157
ANEXO I.....	158
ANEXO II	160

ANEXO III 162

INTRODUÇÃO

A Inteligência Artificial (IA) vive o que Coeckelbergh (2020) denomina de “*hype*”, que pode ser traduzida nessa pesquisa como “moda” ou “propaganda exagerada”. Inserida por vezes invisivelmente nas ferramentas cotidianas do ser humano, a IA é tecnologia que “se esforça para construir entidades inteligentes, bem como entendê-las” (RUSSELL; NORVIG, 2013, p.26, tradução nossa) e, atualmente, vem sendo aplicada às mais diversas áreas do conhecimento, como, por exemplo, na criação dos carros autônomos; no diagnóstico e prevenção das mais diversas doenças; na área de *marketing* digital para modulação de comportamentos; no setor agropecuário e nas artes.

Na área jurídica não é diferente. Existem inúmeras formas de aplicação da IA ao Direito, principalmente pelo fato de que “certas atividades dos juristas são potencial e parcialmente automatizáveis” (SURDEN, 2014, p. 103).

Atualmente, podem-se exemplificar alguns resultados importantes na criação de modelos preditivos que, por aprendizagem com casos passados, preveem decisões judiciais e possíveis resultados da ação, tempo de duração do processo e outros aspectos como valor de condenação (ALARIE; NIBLETT; YOON, 2018)¹. Na automação, classificação e gestão de documentos, os resultados com uso de aprendizado de máquina permitem identificação de documentos em categorias e agrupamento em certas classes (SURDEN, 2014). Na gestão de escritórios e departamentos jurídicos (CESSNA, 2015) e na Resolução *online* de conflitos e criação de *bots* (robôs que simulam comportamento humano e conseguem interagir com as pessoas, por exemplo, respondendo a perguntas) (WANG, 2009; RULE, 2002), assim como na automação de peças e contratos (ALARIE; NIBLETT; YOON, 2018).

Embora sejam diversas as formas atuais de aplicação, inúmeras polêmicas sobrevêm à tona quando se decide aplicar IA ao Direito. De um lado, podem-se citar diversos benefícios como, por exemplo, o estímulo à construção de amplo banco de dados utilizados para criação algorítmica, que servirão para diagnosticar as mais

¹ Os modelos preditivos também são chamados por alguns autores de jurimetria, por utilizarem métodos estatísticos para medir o histórico decisório de magistrados. Disponível em: <<https://www.conjur.com.br/2018-out-12/fred-ferraz-jurimetria-ferramenta-importante-direito>>. Acesso em: 20 mar. 2020.

diversas questões e que podem auxiliar na melhoria da administração da justiça; o tratamento equânime dos processos que chegam ao Poder Judiciário; a melhora na gestão de processos e redução de gastos financeiros e temporais; a melhoria na administração de escritórios, a redução de casos judicializados etc. De outro, há críticas que podem ser feitas e que se relacionam aos graves riscos que a IA pode trazer, tais como: o reforço de vieses discriminatórios, a opacidade algorítmica que pode ser incompatível com a necessidade de fundamentação das decisões judiciais e o alto risco criado pelo uso de dados processuais pessoais, além dos enormes desafios para formação dos novos profissionais dos cursos jurídicos (MAIA FILHO; JUNQUILHO, 2018).

Dessa forma, não obstante os inúmeros avanços que a aplicação da IA ao Direito pode trazer, é fundamental compreender que existem também externalidades éticas negativas, como a opacidade (PASQUALE, 2015), a relativização da privacidade (ZUBOFF, 2019) e os vieses discriminatórios (O'NEIL, 2016), conceitos que serão melhor explorados na pesquisa. Tais questões fazem surgir a necessidade premente de regulação ética em qualquer projeto de desenvolvimento e aplicação de IA (HOVEN; VARMAAS; POEL, 2015).

A despeito dessas preocupações, já se podem citar exemplos práticos de aplicação de IA ao Direito no Brasil e no mundo. A criação do *Ross* foi uma das primeiras iniciativas desenvolvidas utilizando-se do sistema Watson pela IBM. *Ross* é um robô que realiza buscas jurisprudenciais e apresenta respostas aos problemas jurídicos que lhe são apresentados, o que já levou a substituição de diversos advogados em escritórios dos Estados Unidos da América (EUA) (GIUFFRIDA; LEDERER; VERMEYS, 2018).

No Brasil, o Judiciário tem realizado significativos investimentos em programas que empregam IA como ferramenta, que de alguma forma ajudem na gestão processual e eficiência da prestação jurisdicional. Pode-se citar, a título de exemplo, o Projeto Victor, desenvolvido no Supremo Tribunal Federal (STF); o Projeto Sócrates do Superior Tribunal de Justiça (STJ); a ferramenta RADAR desenvolvida no Tribunal de Justiça de Minas Gerais (TJMG); a articulação do Tribunal de Justiça de São Paulo (TJSP) para lançamento de projeto em IA. Na iniciativa privada o país também tem

mais de uma centena de *lawtechs* e *legaltechs*², que apresentam variados serviços jurídicos³.

Em outros países, como nos EUA, o Estado de Wisconsin utiliza a polêmica ferramenta algorítmica denominada *COMPAS* (*Correctional Offender Management Profiling for Alternative Sanctions*) que realiza sugestões aos juízes para que esses atribuam pena e regime de seu cumprimento ao réu, sem, no entanto, explicar as razões pelas quais essas sugestões são feitas.

Na União Europeia (UE) cabe ressaltar que a experiência diverge da brasileira. Isso porque a Comissão Europeia para Eficiência da Justiça (CEPEJ), antes mesmo do desenvolvimento de projetos de IA aplicados ao Direito, já possuía diversas diretrizes sobre a ética na IA, proteção de dados, dentre outros⁴, o que demonstra preocupação da UE com questões regulatórias⁵. Tal fator justifica que a pesquisa observe como UE e Brasil estão lidando com a ética na IA empregada no Direito.

A prudência determina a preocupação com o controle do manejo da IA e com o desenvolvimento de mecanismos éticos que respondam aos desafios de humanização impostos pelas tecnologias chamadas inteligentes. A temática tem aflorado no mundo e fez surgir amplas pesquisas no que se tem chamado de *Explainable AI* (*XAI*), as quais exploram e pretendem avançar na explicação das decisões tomadas pela IA, em especial as que utilizam os modelos de redes neurais, cujos resultados são per se opacos (BARREDO ARRIETA et al., 2020; CERVANTES et al., 2019; RUSSELL, 2019). A

² “Enquanto a *legaltech* visa melhorar a eficiência dos escritórios e departamentos de advocacia, a *lawtech* oferece ferramentas “self-service” para pessoas e negócios”. Disponível em: [/https://fintech.com.br/blog/startup/diferenca-entre-lawtech-legaltech/#:~:text=Afinal%20qual%20a%20diferen%C3%A7a%20entre%20lawtech%20e%20legaltech%3F&text=Enquanto%20a%20legaltech%20visa%20melhorar,service%E2%80%9D%20para%20pessoas%20e%20neg%C3%B3cios](https://fintech.com.br/blog/startup/diferenca-entre-lawtech-legaltech/#:~:text=Afinal%20qual%20a%20diferen%C3%A7a%20entre%20lawtech%20e%20legaltech%3F&text=Enquanto%20a%20legaltech%20visa%20melhorar,service%E2%80%9D%20para%20pessoas%20e%20neg%C3%B3cios). Acesso em: 13 dez. 2021.

³ A Associação Brasileira de Lawtechs e Legaltechs produz um “radar” que apresenta algumas dessas iniciativas privadas. Disponível em: <https://www.ab2l.org.br>. Acesso em: 13 de mar. 2020.

⁴ Para exemplificar citam-se alguns documentos legislativos importantes apontados pelo Laboratório de Políticas Públicas e Internet (LAPIN): “Ethical Guidelines for a Trustworthy AI (EU HLEG-AI, 2019); New Guidelines on Artificial Intelligence and Data Protection (Convenção 108, 2019); Declaration On Ethics And Data Protection In Artificial Intelligence (Global Privacy Assembly, 2018); Towards a Digital Ethics (Comitê de Ética da European Data Protection Supervisor - EDPS, 2018); EU guidelines on ethics in artificial intelligence (European Parliament, 2019); Liability for Artificial Intelligence (European Commission, 2019); Artificial Intelligence: the global landscape of ethics guidelines (ETHZurich, 2019); Policy and investment recommendations for trustworthy Artificial Intelligence (EU HLEG-AI, 2019); AI Policy Sourcebook (EPIC, 2020)”. Disponível em: https://9977a902-e455-46d9-8a7b-0ac71f155f93.filesusr.com/ugd/77388c_3dc8a0975e5542c69c7f6fefdd50ad3b.pdf. Acesso em: 12 mar. 2020.

⁵ No mesmo sentido o Governo francês, em 2019 editou a Lei n. 222, que estabeleceu consequências penais para quem realiza análise estatística para fins preditivos de decisões judiciais. Sobre o tema, conferir interessante análise feita por Sofia Marshallowitz em artigo no Jota <https://www.jota.info/opiniao-e-analise/artigos/o-que-pretende-a-franca-em-proibir-a-jurimetria-18062019>. Acesso em: 22 mai. 2020.

produção científica na área de *XAI* tem aumentado nos últimos anos e se dedicado a estudar técnicas e metodologias para compreensão e inteligibilidade dos modelos de aprendizagem profunda (Vilone & Longo, 2020).

Recentemente o Brasil também tem demonstrado inquietação com o tema da governança no emprego de IA, bem como com a criação de estratégias de desenvolvimento da área. Nesse sentido, o Ministério da Ciência, Tecnologia, Inovações (MCTI) criou pela Portaria MCTI nº 4.617, de 6 de abril de 2021, a Estratégia Brasileira de Inteligência Artificial (EBIA), que dispõe sobre as ações do Brasil em matéria de IA. Muitas críticas, com as quais se concorda nessa pesquisa, foram feitas à EBIA (o autor Ronaldo Lemos denominou a Estratégia brasileira de “patética”⁶), eis que a mesma traz um conceito impreciso de IA e não propõe uma agenda efetiva de investimentos e pesquisas na matéria. Apesar de muito criticada por alguns, representa que de alguma maneira o Estado brasileiro tem se atentado ao assunto.

Além disso, está em tramitação o Projeto de Lei n. 21/20 de autoria do Deputado Eduardo Bismarck (PDT-CE), que pretende criar o marco legal do desenvolvimento e uso da IA pelo poder público, por empresas, entidades diversas e pessoas físicas. O texto, aprovado em regime de urgência⁷ na Câmara dos Deputados e em tramitação no Senado, estabelece princípios, direitos, deveres e instrumentos de governança para a IA⁸.

Tais iniciativas, apesar de demonstrarem certa preocupação do Poder Público com o desenvolvimento da IA no país, ainda se mostram insignificantes em face dos relevantes impactos sociais e desafios éticos que as aplicações dessa tecnologia podem trazer. O recorte que este estudo faz é sobre o controle ético nas diversas possibilidades de aplicação da IA ao ramo do Direito.

Isso porque, mesmo diante do descrito panorama atual e apesar do desenvolvimento e implantação de projetos em IA no Direito brasileiro, em busca

⁶ Em sua coluna no jornal Folha de São Paulo, Ronaldo Lemos apresentou estratégias desenvolvidas por outros países para criticar duramente a EBIA. Disponível em: [https://www1.folha.uol.com.br/colunas/ronaldolemos/2021/04/estrategia-de-ia-brasileira-e-patetica.shtml#:~:text=O%20C3%BAnico%20momento%20em%20que,em%20capacidade%20de%20atrair%20talentos](https://www1.folha.uol.com.br/colunas/ronaldolemos/2021/04/estrategia-de-ia-brasileira-e-patetica.shtml#:~:text=O%20C3%BAnico%20momento%20em%20que,em%20capacidade%20de%20atrair%20talentos.). Acesso em: 06 jan. 2022.

⁷ Cabe aqui a crítica importante de que assunto tão complexo precisa ser amplamente debatido sem urgência na sociedade.

⁸ Mais informações sobre o Projeto podem ser encontradas no site da Câmara. Disponível em: <https://www.camara.leg.br/noticias/641927-projeto-cria-marco-legal-para-uso-de-inteligencia-artificial-no-brasil/>. Acesso em: 10 mar. 2020.

realizada no Portal de Periódicos da Capes e à Rede Virtual de Bibliotecas (RVBI), encontram-se pesquisas que abordam o tema da ética na aplicação geral da IA, mas ainda são escassos no Brasil os trabalhos que abordam o tema dos limites da aplicação ética da IA ao Direito⁹.

Como se pode perceber, ainda há muito que se avançar nas pesquisas sobre a questão, eis que a tecnologia vem sendo implementada no Brasil, sem a necessária compreensão de seus possíveis efeitos e a devida regulação ética.

Por essas razões, a presente pesquisa pretende trazer contribuições que apontam para a aplicação ética da IA ao Direito. Além disso, a atualidade e originalidade do tema, tão pouco explorado no país, pretende proporcionar relevantes avanços nos estudos que vêm sendo desenvolvidos no Brasil e no mundo sobre o assunto.

Diante desse panorama a pesquisa busca resposta à seguinte questão: **que parâmetros éticos são identificados e podem ser considerados no desenvolvimento da IA ao Direito?**

Para tanto, tem como objetivo geral:

. **Analisar os parâmetros essenciais/mínimos, presentes na literatura, nas diretrizes regulatórias e na percepção de especialistas, para controle ético da IA aplicada ao Direito.**

Os objetivos específicos são:

. Descrever o conceito de IA para a pesquisa e as possibilidades de aplicação da IA ao campo do Direito;

. Investigar o perfil das pesquisas e a evolução do tema do controle ético na IA nos artigos publicados na base *Web of Science*, no período de 2017 a 2021;

. Analisar as iniciativas regulatórias da CEPEJ e a Resolução n. 332 do CNJ para controle ético de IA aplicada ao Direito;

. Averiguar a percepção dos limites éticos à IA, dos atores que de algum modo se relacionam ao tema;

. Confrontar os resultados para compreender quais os parâmetros mínimos de controle ético da IA aplicada ao Direito.

⁹ Podem-se citar aqui algumas obras: “Inteligência Artificial e Direito: ética, regulação e responsabilidade”, que é uma coletânea de artigos coordenada por Ana Frazão e Caitlin Mulholland (2019) e a recente obra “Inteligência Artificial e Direito: Convergência Ética e Estratégica” do professor Fabiano Hartmann Peixoto (2020).

A hipótese é de que existem parâmetros éticos mínimos para balizar o controle da IA aplicada ao Direito a serem descobertos após os *inputs* trazidos pela pesquisa.

Para cumprir esses objetivos, no Capítulo 1, apresenta-se a contextualização do cenário atual da IA e aborda-se seu conceito. Em seguida, no Capítulo 2, investigamos o perfil das pesquisas e a evolução do tema do controle ético na IA nos artigos publicados na base *Web of Science*, no período de 2017 a 2021. No Capítulo 3, realiza-se uma análise comparativa, apontando de que forma Brasil e União Europeia têm se preocupado em estabelecer limites éticos para as atividades desenvolvidas por IA no Direito.

No Capítulo 4, para averiguar a compreensão dos limites éticos à IA dos atores que de algum modo se relacionam ao tema, foram entrevistados 11 sujeitos (três profissionais da área de tecnologia e quatro magistrados que atuam em projetos de IA ao Poder Judiciário e quatro pesquisadores das ciências sociais) e apresentadas suas percepções. Por fim, o Capítulo 5 confronta, por meio da triangulação de métodos, os resultados obtidos por cada *input*, para trazer como *output* a compreensão dos parâmetros mínimos de controle ético da IA aplicada ao Direito.

Ao fim, apresentam-se as preocupações éticas e princípios mínimos apontados pela literatura, pelas diretrizes e pelos atores, assim como algumas questões sensíveis em comum que apareceram no debate realizado na área, como as análises preditivas que usam IA no Direito e a possibilidade ou não de IA aplicada ao Direito Penal.

1 INTELIGÊNCIA ARTIFICIAL: CONFUSÃO DO TERMO E EXIGÊNCIAS ÉTICAS

As máquinas autônomas e os robôs superinteligentes geram curiosidade humana há anos na ficção científica. O escritor russo Isaac Asimov, por exemplo, escreveu várias obras sobre o tema e criou em 1950, no livro “Eu, Robô” as três leis da robótica¹⁰, que são lembradas e debatidas ainda hoje. Atualmente a Inteligência Artificial é assunto que está em pauta em muitas discussões acadêmicas havendo, entretanto, muita confusão que paira sobre o termo. Por essa razão, utilizá-lo pode ser bom, por um lado, pois desperta curiosidade, dá energia para produção de pesquisas e geração de investimentos. Por outro lado, também pode ser apresentado de forma equivocada como mera estratégia para atrair atenção de consumidores, clientes e cidadãos (SUSSKIND, 2019). Mas como surgiu a IA e o que esse conceito de fato significa?

O objetivo desse capítulo é descrever como nasceu a concepção de IA que fundamenta a pesquisa e compreender por que seu uso ético deve ser encarado como direito fundamental.

1.1.1 Compreendendo a história da IA

Historicamente, várias áreas como a matemática, a filosofia, a economia, a engenharia da computação, a neurociência, a psicologia e a linguística contribuíram, ao longo dos anos, com pesquisas fundamentais para o desenvolvimento da IA. Para citar exemplos, as teorias da decisão e dos jogos na economia e as descobertas da neurociência sobre o sistema nervoso e neuronal, foram e são responsáveis por dar importantes aportes na evolução da IA (RUSSELL; NORVIG, 2013)

Entretanto, embora a ficção científica e a filosofia refletissem sobre o tema há muitos anos, foi somente há sete décadas que o seminário “*Proposal for the Dartmouth*

¹⁰ 1ª lei: Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal.

2ª lei: Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a Primeira Lei.

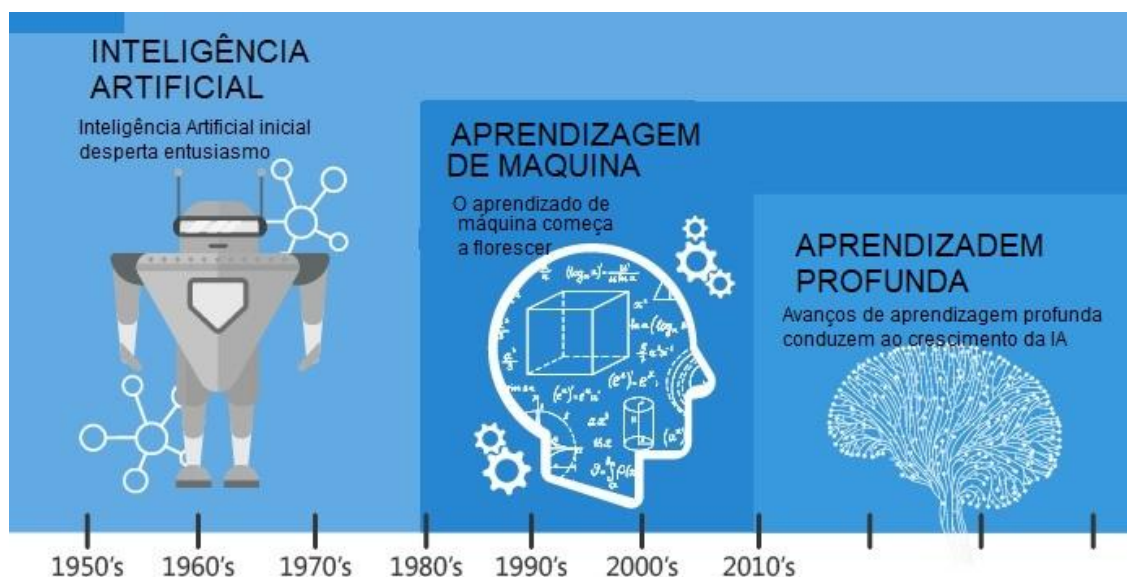
3ª lei: Um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira e Segunda Leis (ASIMOV, 1950).

Summer Research Project on Artificial Intelligence”, ocorrido no verão de 1955, organizado por John McCarthy, apresentou grandes nomes da pesquisa em IA, como Marvin Minsky, Nathaniel Rochester e Claude Shannon. A proposta exibida no seminário trazia pela primeira vez o termo IA e, também, a primeira definição desse novo ramo:

“[...] para a presente proposta o problema da inteligência artificial é fazer uma máquina se comportar de forma a ser chamada de inteligente como se a ação estivesse sendo praticada por um humano” (MCCARTHY et al., 2006, tradução nossa).

Mesmo após mais de setenta anos de pesquisas em IA, quem pretende conceituá-la com precisão ainda encontra dificuldades. Afinal, a história da IA não é de ascensão retilínea, como se pode observar na figura abaixo:

Figura 1- Linha do tempo IA



Fonte: Disponível em: <https://flammarion.wordpress.com/tag/inovacao/>. Acesso em: 04 jun. 2021. (Tradução nossa).

A IA passou por períodos de “inverno”, nos quais eram poucas as pesquisas, produções acadêmicas e descobertas na área. Este é um campo que ainda não tem uma teoria firme, mas vem avançando em seus modelos empíricos principalmente a partir de 2010/2011, quando os “algoritmos de *machine learning* (ml, subárea da ciência da computação) e os de *deep learning* (dl, subárea da ml) receberam impulso de novos avanços, em *hardware* e *software*” (ARBIX, 2020, p. 402). Vive atualmente, portanto, um de seus “verões”, período de muitos investimentos e de importantes resultados.

Dentre os fatores que propiciaram a ascensão dos estudos sobre aprendizagem profunda, estão (a) crescente disponibilidade de grande quantidade de dados (*big data*), (b) maior capacidade computacional e (c) evolução dos algoritmos (KAUFMAN, 2018, p.24).

Mas, “afinal o que é um algoritmo? No seu nível mais fundamental, um algoritmo nada mais é que uma precisa série de instruções para executarem uma tarefa concreta” (KEARNS; ROTH, 2020, p.4). Isso é, um conjunto de códigos que permitem a efetivação de alguma operação.

Além disso, o “ABC das tecnologias de informação e comunicação (*analytics + big data + cloud computing*)” (MAGRANI, 2018, p. 25), caracteriza o contexto em que se desenvolveu a análise de grande volume de dados, aliada ao grande potencial de armazenamento dos mesmos.

É a denominada Quarta Revolução Industrial, por meio da qual a tecnologia domina os meios de produção, provocando intensas transformações na sociedade e na economia. A Quarta Revolução Industrial caracteriza-se pela convergência de tecnologias digitais, dentre as quais a IA tem preponderância e destaque pela capacidade de lidar com grande conjunto de dados e com base neles prever como e quando ocorrerá determinado resultado com alto grau de assertividade (predição). O desenvolvimento da IA e a automatização de processos decisórios provocados pelas técnicas de aprendizado de máquina cumpre papel central nesse contexto (VERONESE; SILVEIRA; LEMOS, 2019; XAVIER; SPALER, 2019). O digital traz à tona a vivência do *homo digitalis*, assim chamado por Han (2019), para denominar a forma como o ser humano atual lida e interage com o mundo à sua volta. Nessa sociedade, “todo clique que eu faço é salvo. Todo passo que eu faço é rastreável. Deixamos rastros digitais em todo lugar. [...] A sociedade digital apresenta uma estrutura especial panóptica” (HAN, 2019, p.122).

Nesse cenário, os indivíduos tendem a estar sempre conectados à internet e, portanto, produzindo dados por meio dos mais variados objetos inteligentes do cotidiano que se comunicam (relógio, televisão, geladeira, carro, celular etc.), fenômeno chamado de internet das coisas (*IoT – Internet of Things*). “A hiperconectividade presente na era da *IoT*, gera circunstâncias que fazem o indivíduo estar permanentemente conectado (*always-on*); a possibilidade de estar prontamente acessível (*readily accessible*); a riqueza de informações; a interatividade; o armazenamento ininterrupto de dados (*always recording*)” (MAGRANI, 2018, p.21).

Na era da internet das coisas os objetos que servem às pessoas conectam-se permanentemente à internet. A máquina, conectada vinte e quatro horas por dia à internet, serve ao indivíduo e relaciona-se com o ambiente por meio de conexões permanentes, capturando, assim, dados que em sua maioria são pessoais, isso é, podem identificar de alguma forma os cidadãos que as utilizam.

A capacidade computacional de armazenamento desse grande volume de dados produzidos a todo instante pelos seres humanos, possibilitou o *boom* das técnicas de aprendizado de máquina “para resolver problemas cada vez mais complexos” (LOURIDAS; EBERT, 2016, p. 110, tradução nossa)¹¹. A proeminência de produção, captação, conservação e guarda de dados em larga escala na era do *big data*, trouxe a “onda de novas demandas nos algoritmos de aprendizado de máquina” (JORDAN; MITCHELL, 2015, p. 257). O *big data* e o aumento da capacidade computacional foram pontos relevantes que possibilitaram a concretização do modelo de redes neurais, ao reduzir a velocidade do processamento e aumentar a oferta de dados quantitativa e qualitativamente.

Tais fatores podem ser apontados como concorrentes para o sucesso do desenvolvimento de modelos algorítmicos de IA, que dependem de muito poder de “processamento e grandes quantidades de dados para serem assertivos em tarefas de classificação de textos, de reconhecimento de faces, de reconhecimento de voz, de predições e de agrupamentos de informações” (SILVA, 2019, p.43).

O movimento da IA, portanto, não é retilíneo, tendo sofrido invernos e períodos de relativo desprestígio. A partir de 2010/2011, contribuíram decisivamente para sua definitiva ascensão a era dourada dos dados (*Big Data*) e o aumento na capacidade de armazenamento desses, somado ao aperfeiçoamento dos algoritmos, sequência de códigos que automatizam decisões, fatores que proporcionaram o desenvolvimento da aprendizagem profunda. Por ser um campo de descobertas relativamente recentes, falta à IA ainda uma robusta teoria que a conceitue com precisão.

1.1.2 Pressuposto: o que é IA para essa pesquisa?

¹¹ Essa capacidade tende a aumentar ainda mais com o desenvolvimento da chamada computação quântica, que aumentará a capacidade de processamento de dados e fará com que a IA desenvolva mais rapidamente (TAYLOR, 2020).

Inicialmente é preciso esclarecer que há críticas ao termo “Inteligência Artificial”, as quais citam principalmente o fato de que a expressão é imprecisa e dá a falsa expectativa de que a IA reproduz precisamente o ser humano, como nas ficções científicas. “O termo “inteligência”, para muitos observadores, sugere talvez que os sistemas mais recentes são, em certo sentido, realmente “conscientes”” (SUSSKIND, 2019, p. 265, tradução nossa)¹². Entretanto, embora criticado, esse é o *topoi* (lugar comum) e termo discursivamente usado nas discussões acadêmicas, sendo fundamental precisá-lo como premissa para a pesquisa.

Para compreender o que é uma inteligência artificialmente construída é necessário antes entender o que é inteligência, característica que, em geral, é atribuída de forma exclusiva ao ser humano. Por isso, aprender sobre IA é também uma forma de autoconhecimento e de auto compreensão da raça humana (RUSSELL; NORVIG, 2013). Todavia, tampouco há consenso sobre o que é o pensamento inteligente. O que seria a inteligência? Capacidade de raciocínio lógico, de compreensão do mundo e de si, de conhecimento emocional, de criatividade, de aprendizado, de resolução de problemas etc.?

Em primeiro lugar, é necessário compreender a IA como substantivo não contável (é impreciso dizer, por exemplo, que foi desenvolvida “**uma** IA”), pois os estudos em IA estão inseridos na ciência da computação e tal como na matemática, na biologia, dentre outros campos, comportam diversos métodos de resolução de problemas específicos¹³. Não existe, tampouco, “uma” IA geral ou forte que imite completamente o comportamento cognitivo do humano em todas as suas nuances (ou que o supera – superinteligência/ultrainteligência), criação presente nos relatos de ficção científica. Ainda estamos distantes da ““IA geral”, a tecnologia para todos os fins que pode fazer o mesmo que o humano é capaz” (LEE, 2019, p. 23).

O termo IA tem sido usado comumente para representar a “IA fraca ou estreita”, a qual abrange abordagens de resolução de problemas específicos. Nesse sentido, Garwood (2018, p. 10, tradução nossa), alerta que há inúmeras formas de conceituar IA, algumas publicadas em livros inteiros dedicados ao assunto que são extremamente

¹² Segundo Susskind (2019, p. 265, tradução nossa), “no jargão filosófico da IA, um sistema que é consciente seria uma exibição do que é conhecido como “IA forte””. A “IA forte”, entretanto, não é objeto dessa pesquisa pois, como se abordará mais a frente, ainda não é realidade no desenvolvimento da IA.

¹³ Disponível em: <https://course.elementsofai.com/1/1>. Acesso em: 30 abr. 2020.

técnicas, outras confusas. O manual de IA dos especialistas Russel e Norvig (2013) define IA como campo que estuda agentes inteligentes.

Segundo Cortiz (2020) “a IA pode ser entendida como um grande guarda-chuva que tem duas grandes abordagens: uma baseada em conhecimento e o aprendizado estatístico”¹⁴. Os agentes baseados em conhecimento “são capazes de aceitar novas tarefas sob a forma de metas descritas de modo explícito, podem alcançar competência rapidamente ao serem informados ou adquirirem novos conhecimentos [...]” (RUSSELL; NORVIG, 2013, p. 349).

Essa abordagem compreende que “todo comportamento inteligente pode ser captado por um sistema que raciocine logicamente a partir de um conjunto de fatos e regras que descrevem o domínio”, isso é, que por meio de sistemas lógicos apreende a compreensão dos especialistas da área de determinado problema a ser solucionado. Atualmente essa abordagem é conhecida como GOF AI (*Good Old-Fashioned Artificial Intelligence*, ou a boa velha IA/a IA tradicional/à moda antiga) (RUSSELL; NORVIG, 2013, 827).

Ensinar computadores a pensar por meio da codificação de regras lógicas (se A, então B), era a tarefa dos desenvolvedores do campo baseado em regras ou sistemas simbólicos/especialistas (LEE, 2019). Já o aprendizado estatístico ou aprendizado de máquina (AM) (*machine learning*, termo cunhado por Arthur Lee Samuel em 1959), compreende a investigação e o estudo e concepção de algoritmos capazes de assimilar e prever dados.

As pesquisas em AM despertaram o surgimento de cinco escolas de pensamento: a) os simbolistas, para os quais o manuseio de símbolos é o que importa para reproduzir a inteligência, que trabalham com a dedução inversa (encontrar algoritmo capaz de generalizar uma dedução); b) os conexionistas, que compreendem o aprendizado inteligente como função do cérebro e utilizam *backpropagation* para reproduzir as conexões neuronais; c) os evolucionários, que entendem que o aprendizado vem da seleção natural e, por isso, usam a programação genética como norte para criação algorítmica; d) os bayesianos, para os quais todo aprendizado a ser reproduzido é incerto e por isso usam algoritmos de inferência probalística e adotam a teoria de Bayes e seus consequentes e os e) analogistas, que acreditam que o aprendizado se dá por

¹⁴ O professor Diogo Cortiz cita essas abordagens em seu curso sobre IA. Disponível em: <<https://www.youtube.com/watch?v=Ze-Q6ZNWpco&t=1136s>>. Acesso em: 02 mai. 2020.

identificação de semelhança entre os acontecimentos e sua reprodução, que empregam máquina de vetores de suporte (DOMINGOS, 2017).

O AM comporta diferentes tipos de abordagens, dentre eles o supervisionado, o não-supervisionado e o por reforço¹⁵. Cada uma delas tem um conjunto de técnicas e é adequada para resolver determinados tipos de situações.

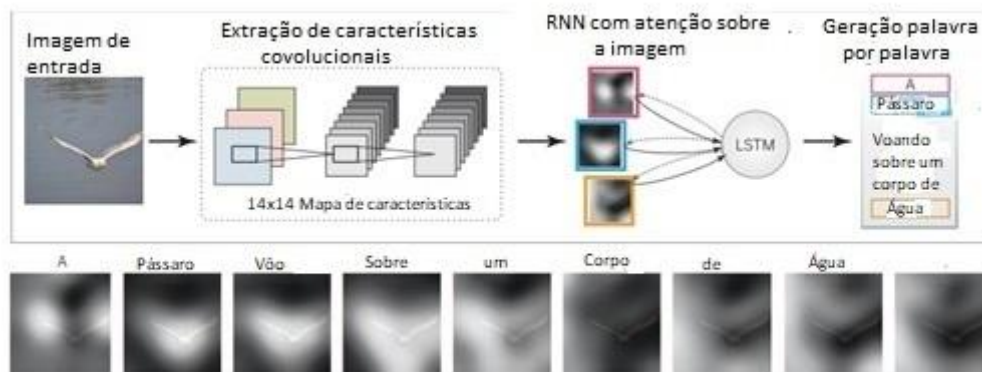
Pelo aprendizado supervisionado, sabendo-se do resultado desejado, o conjunto grande de dados inseridos para criação do modelo são rotulados previamente com a resposta correta necessária para a resolução do problema. Isso é, “o conjunto de treinamento contém dados e a saída correta da tarefa com esses dados” (LOURIDAS; EBERT, 2016, p.110, tradução nossa). Por exemplo, quando o modelo a ser construído objetiva criar algoritmo que separe os *e-mails* que são *spam* daqueles que não são e o conjunto de dados (*dataset*) apresenta previamente exemplos classificados do que é e o que não é *spam*. O aprendizado supervisionado se vale das técnicas de regressão (que inclui predição, árvores de decisão, redes bayesianas, regressão linear, dentre outras) e classificação (que abrange regressão lógica, redes neurais artificiais, dentre outras).

Uma das áreas que mais evoluiu pelo aprendizado supervisionado foi o da aprendizagem profunda (*Deep Learning* (DL)) que faz “uso de algoritmos de otimização baseados em gradiente para adaptar os parâmetros por meio de uma rede de múltiplas camadas baseadas em erro da saída (*output*)” (JORDAN; MITCHELL, 2015, p. 257, tradução nossa). Na Figura 2 demonstra-se o funcionamento de uma rede neural convolucional treinada para interpretar imagens, nela o *output* (saída) é utilizado

[...] por uma rede neural recorrente treinada para gerar uma legenda (em cima). A sequência na parte inferior mostra o foco palavra por palavra da rede em diferentes partes da imagem de entrada enquanto gera a legenda palavra por palavra. (JORDAN; MITCHELL, 2015, p. 257, tradução nossa)

Figura 2- Geração automática de legendas de texto para imagens com redes profundas

¹⁵ Existe também um aprendizado “intermediário” denominado semi-supervisionado por meio do qual a máquina “aprende com uma pequena quantidade de dados rotulados e com uma grande quantidade de dados não rotulados” (HAO; ZHANG, 2016, p. 426).



Fonte: JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, v. 349, n. 6245, 2015 (Tradução nossa)

O AM inclui, portanto, a aprendizagem profunda, que se trata de pilhas de multicamadas de modelos sujeitos a aprendizado, o qual também se desenvolveu no tipo de aprendizado não-supervisionado (LECUN; BENGIO; HINTON, 2015).

[...] Embora grande parte do sucesso prático da aprendizagem profunda venha de métodos de aprendizagem supervisionada para descobrir essas representações, também foram feitos esforços para desenvolver algoritmos de aprendizagem profunda que descubram representações úteis da entrada sem a necessidade de dados de treinamento rotulados. (JORDAN; MITCHELL, 2015, p. 258, tradução nossa)

Por aprendizado não-supervisionado compreende-se aquele que envolve a análise de conjuntos de dados (*dataset*) não rotulados previamente. Esse pode valer-se de técnicas para segmentação (que compreende técnicas como algoritmos genéticos, por exemplo) e para redução de dimensão (que utiliza técnicas como o aprendizado múltiplo, análise fatorial, projeções aleatórias, dentre outras).

A Figura 3 resume as duas abordagens de AM e técnicas que podem ser usadas.

Figura 3- Abordagens de Aprendizado de Máquina e técnicas.



Fonte: LOURIDAS, P.; EBERT, C. Machine Learning. **IEEE SOFTWARE**, 2016 (Tradução nossa)

Ambos os paradigmas são usados quando se está diante de uma grande massa de dados, o que não é necessário na aprendizagem por reforço. A aprendizagem por reforço, inspirada principalmente na psicologia comportamental e neurociência, trabalha com sistema de recompensas para acertos da aprendizagem da máquina. Esse tipo de aprendizado é feito em ambientes controlados que permitam que o sistema aprenda por tentativa e erro (JORDAN; MITCHELL, 2015). Em resumo, na aprendizagem por reforço o algoritmo “aprende como fazer ações através de observações. Cada ação tem um impacto no meio ambiente. O aprendizado por reforço faz julgamentos com base no *feedback* do ambiente observado” (HAO; ZHANG, 2016, p. 426, tradução nossa).

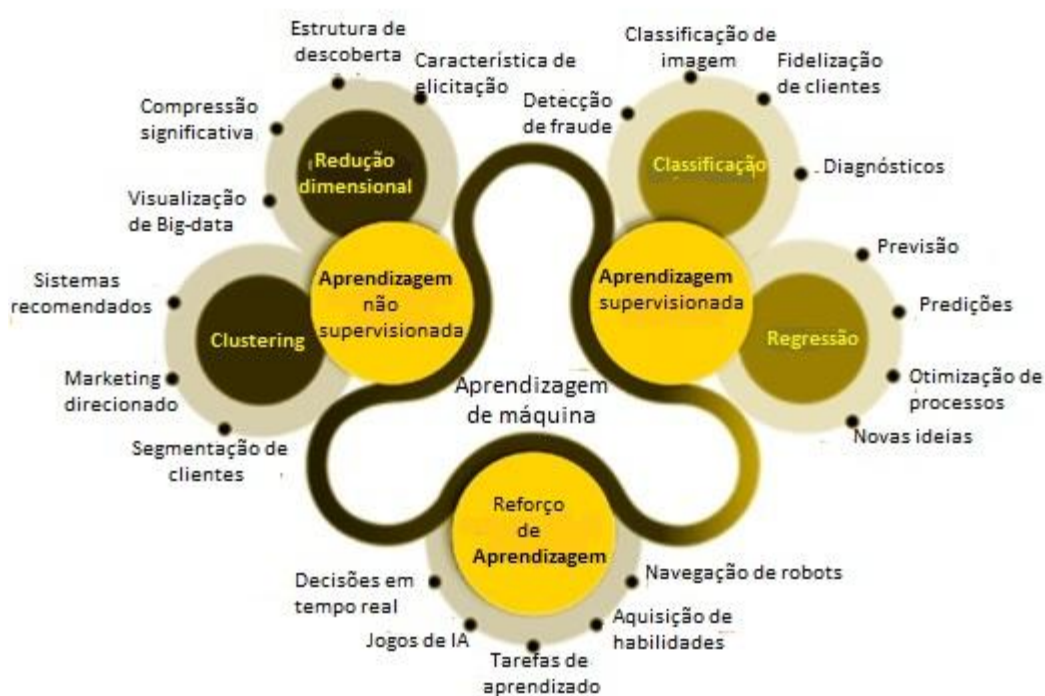
A IA geral, conforme foi afirmado, não existe. As abordagens de IA fraca ou estreita resolvem problemas específicos. Conclui-se que no AM, “cada método de aprendizado, usado separadamente, tem aplicações limitadas, pois requer tipos especiais de *input* e conhecimento e aprende um tipo específico de informação” (TECUCI, 2012, p. 175).

Cada técnica de abordagem se adéqua melhor para desenvolvimento de uma tarefa particular. O aprendizado supervisionado pode ser usado, por exemplo, para desenvolver sistemas antifraude, reconhecimento de imagem e de fala, tradução de idiomas, no desenvolvimento de filtros de *spam* e de sistemas de recomendação (predição). O aprendizado não-supervisionado para pontuação de crédito (*credit score*), em detecção de fraudes, segmentação de clientes, dentre outros. E o aprendizado por reforço pode ser usado para criação de jogos e robótica (DOMINGOS, 2012).

LeCun, Bengio e Hinton (2015, p. 100, tradução nossa) apresentam exemplos de aplicação do aprendizado profundo que possibilitaram o *hype* atual da IA: para manuseio de *big data* em marketing de vendas e promoções; reconhecimento de objetos e padrões para encontro de “pontos fracos”, previsão de fluxos e vídeo para direção autônoma, avaliação e análise de imagens para identificação de formas distintas (como em imagens médicas ou reconhecimento facial e impressão digital) e “heurísticas de segurança que soltam ataques padrão para proteger, por exemplo, portas ou redes”.

Na Figura 4 pode-se verificar um resumo das abordagens de AM e usos que podem ser feitos, por técnica.

Figura 4– Aprendizado de Máquina e utilizações por técnica.



Fonte: Disponível em: <<https://pt.the-ai-academy.com/machine-learning-bootcamp>>. Acesso em: 08 mai. 2020 (Tradução nossa)

De forma resumida, é possível observar que a moderna compreensão sobre IA envolve entendê-la como algo que pode resolver problemas específicos. Um problema de aprendizagem pode ser definido como o problema de melhorar alguma medida de desempenho ao executar alguma tarefa, através de alguns tipos de experiência de treinamento (JORDAN; MITCHELL, 2015, p. 255).

Note-se que, longe da compreensão que a princípio o termo pode nos levar, a IA não é a reprodução exata do comportamento humano, mas apenas algumas funções específicas do cérebro humano. Defini-la passa por compreender como funcionam os modelos de aprendizagem da máquina e saber identificar qual a melhor técnica se ajusta ao problema a ser resolvido. A pesquisa dá enfoque ao modelo de aprendizado de máquina e à técnica de redes neurais artificiais, que é atualmente a mais utilizada.

1.2 POR QUE É (DIREITO) FUNDAMENTAL EXIGIR APLICAÇÃO ÉTICA DA IA?

A compreensão de IA para essa pesquisa foca, portanto, na aprendizagem profunda, modelo que vem sendo mais utilizado e que obteve maiores avanços recentes, pois permite o desenvolvimento de soluções tecnológicas que proporcionam progressos nas mais variadas esferas da vida humana¹⁶. Existem numerosos exemplos de aplicações de sucesso de IA, que colaboram tanto no desenvolvimento de políticas públicas, quanto na otimização de processos e no crescimento exponencial do setor privado. Nesse sentido, inúmeras experiências podem ser apontadas em alguns campos que experimentam profundas transformações¹⁷.

No setor de transportes os maiores exemplos são os veículos autônomos, as tecnologias de IA incorporadas aos automóveis¹⁸ e as plataformas que traçam melhores rotas em tempo real (GORDON-MURNANE, 2018). As mudanças trazidas podem refletir na diminuição do índice de acidentes e danos, no aumento de índices de qualidade e expectativa de vida, na melhoria do trânsito e redução de tempo gasto em deslocamento (SOUZA; OLIVEIRA, 2019).

Na área da segurança também podem ser apontados alguns importantes resultados alcançados pelas pesquisas em IA. Exemplos de aplicações são: *Drones* de rastreamento¹⁹, armas autônomas para fins militares (PEREZ et al., 2017) e uso de

¹⁶ Disponível em: <<https://blogs.scientificamerican.com/observations/ethics-in-the-age-of-artificial-intelligence/>>. Acesso em: 13 maio 2020.

¹⁷ No site da Comissão Norte-Americana de proteção aos consumidores, é possível encontrar vários exemplos de aplicações atuais bem sucedidas de AM. Disponível em: <<https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms>> Acesso em: 20 mai. 2020.

¹⁸ Algumas descobertas interessantes se encontram na seguinte reportagem: <https://www.nytimes.com/2018/01/04/technology/self-driving-cars-aurora.html>. Acesso em: 13 maio 2020.

¹⁹ Uma reportagem do jornal The New York Times anuncia com preocupação a criação de Drone capaz de seguir usuários.

reconhecimento facial para detecção de criminosos²⁰ (LAUTERBACH, 2019). Quem defende seu uso no setor de segurança, aponta como benefícios “redução de custos, preservação da vida humana, melhor precisão.” (SOUZA; OLIVEIRA, 2019, p. 69). Cabe ressaltar, entretanto, que essa pesquisa concorda com as críticas realizadas por autores como Ruha (2021), que são contrários à utilização da IA para reconhecimento facial na segurança pública tendo em vista o potencial discriminatório e de reforço ao encarceramento de determinados grupos sociais, à seletividade penal e à hipervigilância.

Nos setores consumerista, de seguros e financeiro, a IA tem sido usada tanto para ofertar e recomendar produtos e serviços adequados às necessidades do usuário (GORDON-MURNANE, 2018), quanto para elaboração de pontuação de crédito (*credit-score*), para análise de concessão de benefícios e seguros e na detecção de fraudes, por bancos e instituições financeiras (Sadatrasoul et al., 2013). Tais análises permitem a redução de riscos e aumento da rentabilidade para empresas, além de aprimorar a experiência das necessidades de clientes (TZIRULNIK; BOAVENTURA, 2019). Na agricultura, há avanços de AM para monitoramento de campo, controle de pragas, análise de imagem por satélite, robotização da produção e colheita, que aumenta a eficiência no setor alimentício (LAUTERBACH, 2019).

Na área da educação, algumas aplicações são: sistemas tutoriais inteligentes, formas de pesquisas avançadas e detalhadas de referenciais teóricos e aprendizagem personalizada (KAY, 2012). Tais utilizações transformam a experiência educativa, possibilitando maior acesso à informação, melhora na qualidade de pesquisas e redução de custos.

Nas artes, há desenvolvimento de inúmeras tecnologias de AM e aprendizagem profunda no mercado audiovisual para, por exemplo, reconhecimento e identificação de sons e imagens e a reprodução e criação de obras de arte (pinturas, músicas, dentre outros). Com esses avanços, um novo mercado produtivo de artes surge, além de novas formas de criação e reprodução cultural (MIRANDA; WILLIAMS, 2015; ZICKLER et al., 2013; QIAN et al., 2017).

Há aplicações de aprendizagem profunda também na área de *Games* que criam realidades virtuais (MORAV et al., 2017; LAWTON, 2011) e em redes sociais e interações por *chatbots* (MOU; XU, 2017). As redes sociais são atualmente uma das

²⁰ Sistemas de reconhecimento facial vêm sendo usados ao redor do mundo, principalmente com fins de colaborar com a segurança. No Brasil, inclusive, a Bahia foi o primeiro Estado a usar essa tecnologia. Disponível em: <<https://www.correio24horas.com.br/noticia/nid/entenda-como-funciona-o-reconhecimento-facial-que-ajudou-a-prender-mais-de-100-na-ba/>>. Acesso em: 14 maio 2020.

maiores fontes de captação de dados para desenvolvimento de técnicas em aprendizado de máquina que propiciem avanços na interação humana (MAGRANI, 2019). E, para além disso, constituem a base para os novos modelos de negócio da indústria 4.0, das *big techs* e das diversas plataformas tecnológicas.

Na saúde, a aplicação de técnicas de AM na medicina tem servido para diagnóstico de doenças por imagem, como, por exemplo, o câncer de pele (ZHANG ET AL., 2017) e avaliação preditiva de métricas de propensão a patologias (PARK; HAN, 2018). Em resumo, técnicas que atuam na “prevenção, no diagnóstico ou no tratamento das condições do corpo e da mente” (SOUZA; OLIVEIRA, 2019, p. 67).

Ainda na área da saúde, necessário citar os importantes avanços da IA que têm ajudado na atual conjuntura da pandemia mundial do COVID-19. Recente pesquisa encontrou sete aplicações importantes da IA nesse contexto: 1) Detecção e diagnóstico precoce da infecção; 2) Monitoramento do tratamento; 3) Rastreamento de contato dos indivíduos; 4) Projeção de casos e mortalidade; 5) Desenvolvimento de medicamentos e vacinas; 6) Redução da carga de trabalho dos profissionais de saúde; 7) Prevenção da doença (VAISHYA ET AL., 2020). Além disso, podem-se citar as soluções chamadas de *contact tracing*, por meio do qual rastreiam-se informações dos indivíduos e permite-se que o cidadão seja alertado do contato com alguém que teve a doença (CHEN ET AL., 2019).

A conexão dos diversos setores que utilizam IA pode colaborar com a criação das chamadas “Cidades Inteligentes” (*Smart Cities*), desenhadas com modelos tecnológicos inovadores e que adotam inúmeras soluções para o desenvolvimento socioeconômico sustentável (VODĀ & RADU, 2015). É o caso do protótipo de cidade inteligente de Songdo, na Coreia do Sul, cuja construção se iniciou em 2003 e que usa tecnologias para gestão do lixo, mobilidade urbana, controle ambiental, sistema de saúde etc. (ALTENHOFER ET AL., 2016). As soluções tecnológicas adotadas melhoram o planejamento urbano e otimizam a gestão das cidades.

Como se pode observar, o resultado da aplicação de IA, que atualmente se dá nas mais diversas atividades humanas, impacta profundamente nesses nichos e requer um rearranjo nas formas de pensar as soluções, tendo em vista as consequências que sua utilização pode trazer. Afinal, a despeito dos diversos benefícios e avanços enumerados, que significam maior eficiência na resolução de problemas os mais variados, é necessário assegurar que essa aplicação se dê de forma ética.

Além disso, o fato de a definição clássica de IA compreendê-la como um produto que ganha certa autonomia de aprendizagem em relação ao humano criador, coloca o desafio ético de controle desses agentes “inteligentes” (FLORIDI, 2019). Diante das evoluções que a IA tem trazido em diversas aplicações práticas do cotidiano, é hora de “assegurar que as máquinas se manterão benéficas aos seres humanos para sempre” (RUSSELL, 2019, p. xii).

Acredita-se que “de modo ideal, a ética é um código de leis que prescreve o comportamento “universalmente” correto, isto é, para todas as pessoas em todos os momentos” (BAUMAN 2011, p. 22)²¹. A noção de ética que permeia a pesquisa, portanto, diz respeito à preocupação de cada indivíduo, ao tomar uma decisão, em fazer a escolha acertada, com base em determinados valores que subjazem cada uma de suas decisões cotidianas.

[...] As questões éticas dizem respeito a julgamentos de certo e errado, bom e ruim, bem como questões de justiça, virtude e responsabilidade social. Portanto, definimos aqui o raciocínio ético como a capacidade de identificar, avaliar e desenvolver argumentos éticos a partir de uma variedade de posições (DIGNUM, 2019, p. 35, tradução nossa)

O desafio de refletir sobre os possíveis efeitos e consequências de determinada inovação é questão fundamental sempre que surge uma nova tecnologia. Considere-se como exemplo a criação da bomba atômica e a preocupação de cientistas - que foi apenas posterior - para restringir seu uso (KEARNS; ROTH, 2020). Conceber ética em IA faz parte do estudo da ética em tecnologia.

[...] A ética tem raízes antigas, refletindo seu papel perene e central na vida humana, enquanto a IA se desenvolveu significativamente apenas no último meio século. E é o crescimento explosivo do poder e do potencial da IA na última década ou mais que trouxe ao público a importância crucial de abordar as muitas questões éticas que ela suscita (BODDINGTON, 2017, p. xi, tradução nossa).

Quando o assunto é IA, o ponto central está em compreender que o aprendizado de máquina produz como resultado indicadores para automatização da tomada de decisões. Nesse sentido, essas decisões precisam ser explicáveis, transparentes, passar

²¹ Pode-se trazer à discussão a tese do objetivismo moral, por meio do qual acredita-se que decisões éticas são tomadas com base em valores que assegurem a crença na correção racional da ação (ATIENZA, 2017). Em sentido contrário, Adeodato (2002) entende que não há racionalidade nos juízos éticos, mas alerta que o ceticismo não significa abdicar de parâmetros éticos. Com o ceticismo busca-se a eudaimonia, a felicidade sem fanatismo por meio da concepção de uma vida simples.

por rigoroso controle e envolver o humano. E, dessa forma, é fundamental pensar em como se pode desenvolvê-la de forma ética.

As pesquisas em IA devem seguir, portanto, a perspectiva de padrões e balizas éticos para seu desenvolvimento. Falar em aplicação ética de IA envolve compreender vários aspectos, dentre eles: a) quais as funções de cada tecnologia desenvolvida, se essas funções são fundamentais e estão sendo cumpridas; b) a metaética, que perpassa a investigação da eficácia das normas criadas para regular a aplicação da IA; c) o grau de normatividade (forte ou fraca) dessas normas (DALY et al., 2019).

Isso é, pesquisar a aplicação ética da IA abrange: entender previamente quais as aplicações e finalidades dessa tecnologia em cada caso de uso para poder averiguar posteriormente se essas utilidades e propósitos estão sendo atingidos; compreender como criar normas eficazes e tecnologicamente possíveis que equilibrem ampla utilização à limitação adequada (sem excessos que impeçam avanços tecnológicos).

O desafio é saber quais devem ser os limites ao uso da IA para compreender como devem ser concebidas as normas que regularão esses usos, bem como para refletir sobre a eficácia dessa necessária normatividade. Isso porque o desenvolvimento de IA, a despeito de trazer inegáveis avanços e vantagens, impõe inúmeros desafios éticos por

[...] trazer alguns novos problemas, por exemplo [...]:
 . As pessoas podem perder seus empregos para a automação;
 . As pessoas poderiam ter muito (ou pouco) tempo de lazer;
 . As pessoas poderiam perder seu sentido de identidade;
 . Sistemas de IA poderiam ser utilizados para fins indesejáveis;
 . O uso de sistemas de IA poderiam resultar na perda de responsabilidade;
 . O sucesso da IA poderia significar o fim da raça humana (RUSSELL;
 NORVIG, 2013)

Uma série de questões surgem e podem ser pontuadas no desenvolvimento de IA, como: o fato de que a IA, ao automatizar determinadas tarefas, pode gerar a substituição do humano pela máquina e motivar desempregos; alguns fins indesejáveis, como são exemplo as armas de destruição em massa automáticas ou a possibilidade de perda substancial da privacidade; a preocupação com a forma de responsabilização das máquinas que executam os projetos de IA; e até mesmo a inquietação extrema que é a criação de máquinas superinteligentes e capazes de extinguir a raça humana.

Recentemente os “impressionantes avanços nas capacidades e aplicações dos sistemas de IA trouxeram oportunidades e riscos da IA para sociedade”, eis que utilizar IA significa transferir total ou parcialmente o poder de decisão de modo voluntário aos

produtos da tecnologia (FLORIDI; COWLS, 2019, p. 7, tradução nossa). E, assim, conforme foi afirmado, permitir que a máquina tome a decisão inteligente pelo ser humano. Pasquale (2015) compreende que é necessário importar-se com as consequências da aplicação de IA, pois o poder e a autoridade vêm se expressando em forma algorítmica.

Ora, a IA é desenvolvida e criada por seres humanos. Isso significa que são os seres humanos que, ao contrário do que se pode imaginar, há muitos aspectos em projetos de IA, sob controle dos humanos como: são humanos que estão inseridos no processo de produção prévia dos dados, na captura e escolha do conjunto/base de dados que será usado em cada projeto, na criação dos algoritmos, na avaliação dos níveis de riscos e qualidade dos resultados, na contratação da equipe desenvolvedora, na definição de financiamento ou não de cada projeto, na escolha e contratação do ente realizador etc.

O relatório “O futuro da IA no sistema judiciário brasileiro”²², produzido pelo Instituto de Tecnologia e Sociedade do Rio de Janeiro, sugere, por exemplo, quatro métricas avaliativas que podem ser utilizadas para os resultados da IA no Judiciário:

Acurácia: com que frequência o modelo está correto

$Acurácia = \frac{Verdadeiros\ positivos + Verdadeiros\ negativos}{Exemplos\ totais}$

Exemplos totais

Precisão: avalia com que frequência a ferramenta está correta quando prevê um resultado positivo

$Precisão = \frac{Verdadeiros\ positivos}{Verdadeiros\ positivos + Falsos\ positivos}$

Verdadeiros positivos + Falsos positivos

Revocação (Recall): Avalia a frequência de falsos negativos

$Recall = \frac{Verdadeiros\ positivos}{Verdadeiros\ positivos + Falsos\ Negativos}$

Verdadeiros positivos + Falsos Negativos

Pontuação F1: Calcula a frequência com que o modelo está correto, incorporando a pontuação de precisão e de recall.

$F1 = 2 \times \frac{Precisão \times Recall}{Precisão + Recall}$ (ITS, 2020).

Por isso, faz sentido a preocupação com sua aplicação ética, pois essa só serve se utilizada para favorecer e trazer os maiores benefícios aos indivíduos. Nesse sentido, segundo Floridi (2016, tradução nossa)²³

²² O relatório completo está disponível em: <https://itsrio.org/wp-content/uploads/2020/07/TRADUC%CC%A7A%CC%830-The-Future-of-AI-in-the-Brazilian-Judicial-System.pdf>. Acesso em: 13 jan. 2022.

[...] Deveríamos tornar a IA favorável ao meio ambiente. [...] Deveríamos fazer a IA ser humana. [...] Deveríamos fazer a estupidez da IA funcionar para a inteligência humana. [...] Devemos fazer com que o poder preditivo da IA trabalhe por liberdade e autonomia. [...] E, finalmente, devemos fazer com que a IA nos torne mais humanos (FLORIDI, 2016, tradução nossa)

É comum que o desenvolvimento de tecnologias carregue consigo o valor da neutralidade científica, isso é, a ideia de que são apenas instrumentos para consecução de fins e que não endossam, desse modo, nenhuma visão de mundo, ideologia ou posicionamento (MORROW, 2014). O bom desempenho das pesquisas em IA para realização de tarefas cognitivas antes realizadas pelos seres humanos, vêm trazendo avanços inegáveis e os resultados alcançados – assim como os de quaisquer tecnologias – estão cobertos pelo manto de uma pretensa neutralidade científica dos números e códigos (EN; PÖLL, 2016; MULHOLLAND; FRAJHOF, 2019). A aura de imparcialidade atinge, portanto, a internet, os governos e as grandes corporações que lidam com IA e com suas aplicações, como a Google, Bing, dentre outras (CRANE, 2012).

Entretanto, ao contrário do que se imagina, a IA não é neutra. Em sentido oposto, como se pôde perceber no capítulo anterior, por sugerir decisões, realizar predições e apontar tendências para humanos, a IA coloca novos desafios no campo ético. Nesse sentido, requer iniciativas legislativas que tracem limites à sua aplicação, a qual não deixa de ser necessária, útil e fundamental quando realizada com prudência.

Pensar a aplicação ética de IA é entender que “a IA não tem uma ética própria, trata-se de elaborar um conjunto de melhores práticas que possa ser replicado em uma ampla variedade de configurações” (KAUFMAN, 2020)²⁴. Isso é, considerando a ação e intervenção humana no processo de criação, desenvolvimento e controle dos riscos e resultados das aplicações de IA, é para o humano que se deve pensar a ética em sua consecução. Sem dúvidas a IA impõe novos desafios éticos aos cidadãos e quando se fala em aplicação ética da IA parte-se dessa premissa para compreender novos caminhos de controle da ação humana que resulta nessa nova tecnologia.

²³ FLORIDI, Luciano. Should we be afraid of AI? Disponível em: <<http://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>>. Acesso em: 19 mai. 2020.

²⁴ Discorda-se nesse sentido, de Peixoto (2020, p. 31) que se filia “à ideia de que não há ainda, claramente, um acordo sobre o que seria um conceito de ética em IA”. Isso porque, entende-se como Kaufman (2020), que a ética é da ação humana. Disponível em: <<https://epocanegocios.globo.com/colunas/IAgora/noticia/2020/07/etica-e-objeto-da-acao-humana-nao-existe-etica-da-inteligencia-artificial.html>>. Acesso em: 23 jul. 2020.

Por isso há que se pensar em ética humana no desenho e criação dos produtos de IA, percepção trazida por Dignum (2019, p. 6-7, tradução nossa) que aponta três dimensões éticas a serem respeitadas quando se pensa em uma IA responsável:

- 1) Ética **no design**, isso é, pensar em princípios mínimos de aceitabilidade social para a engenharia e para avaliação dos projetos de IA:

[...] Aqui, o objetivo é garantir que as equipes de desenvolvimento de sistemas de IA estejam atentas às possíveis consequências para indivíduos e sociedades, antecipando as consequências das escolhas de projeto, refletindo sobre o problema a ser resolvido, envolvendo todos os interessados, verificando e validando o projeto e tomando as medidas apropriadas para garantir a aceitabilidade social, legal e ética do sistema. Isso significa que precisamos perceber que os princípios de *accountability*, responsabilidade e transparência (ART) estão no centro do *design* dos sistemas de IA. (DIGNUM, 2019, p. 6-7, tradução nossa).

- 2) Ética **pelo design**, que reflete sobre como reagem e trabalham os sistemas de IA, preocupando-se com

[...] (a) os dados sobre a representação e uso de valores morais pelos sistemas de IA, (b) compreensão do significado e especificação de restrições adequadas ao comportamento do sistema e (c) integração de recursos de raciocínio ético como parte dos algoritmos que determinam o comportamento de sistemas autônomos artificiais (DIGNUM, 2019, p. 6-7, tradução nossa).

- 3) Ética **para design(ers)**, que trata das questões relacionadas aos desenvolvedores dos sistemas de IA e daqueles que utilizam tais criações. Nesse sentido,

[...] refere-se aos códigos de conduta, às exigências regulatórias e aos padrões e processos de certificação que garantem a integridade de todos os atores à medida que pesquisam, projetam, constroem, empregam e gerenciam sistemas artificialmente inteligentes. Isso é para garantir que eles considerem o impacto social de suas escolhas de design e tomem as medidas necessárias para minimizar o impacto negativo e o uso duplo de seus resultados. Isso inclui a aderência a códigos de conduta específicos e a definição e uso de padrões, regulamentos e processos de certificação que garantem a integridade dos desenvolvedores, produtos e serviços (DIGNUM, 2019, p. 6-7, tradução nossa).

Em resumo, o estímulo ao amplo desenvolvimento dos modelos de IA é premente. Mas fomentar e impulsionar a IA não é, por si, suficiente. Assim como qualquer tecnologia, a IA precisa de balizas éticas **no design**, que garantam aceitabilidade social dos modelos criados; **pelo design**, que imponham integridade ética no uso de dados, na compreensão e entendimento da representação e concepção dos modelos e ética **para o design**, com a criação de e respeito a códigos e guias de conduta, certificações de qualidade ética para que os atores envolvidos no processo de desenvolvimento tenham garantias e possam conduzir projetos de acordo com referenciais regulatórios socialmente construídos.

Quando se fala em ética para IA, portanto, é necessário entender que todos os envolvidos na implementação, concepção, elaboração e criação precisam se atentar e devem preocupar-se em seguir um conjunto de práticas, condutas e mecanismos que estejam de acordo com padrões éticos. Mais uma vez, parte-se da premissa de que são muitas as etapas em que o humano participa: desde a escolha do ente desenvolvedor e da respectiva equipe, à gestão e escolha da base de dados, ao desenvolvimento dos algoritmos, ao controle dos riscos, até a aferição da acurácia (ou qualidade dos resultados) ou medição do modelo produzido.

A importância, portanto, está em compreender que para se chegar a IA ética deve-se pensar na responsabilidade de todos aqueles que lidam com a IA, o que significa capacitação das equipes para compreensão dos riscos do produto desenvolvido e busca constante pela minimização de danos ou incorreções. Os atores envolvidos na construção dos modelos de IA podem ser considerados artistas, que precisam massagear o modelo, treinando-o para definir como esse pode impactar no resultado final, avaliando-o, criando o melhor design e ajustando a melhor maneira de mostrar para o cidadão leigo seu uso²⁵.

Compreender o design ético da aplicação de IA, como chave para o debate sobre ética e IA, é perceber que embora existam alguns aspectos que saem do controle humano (como é a opacidade dos modelos de redes neurais), ainda assim, é o humano quem conduz muitas etapas do que é concebido em modelos algorítmicos.

Além disso, se a IA é uma das tecnologias mais importantes da atualidade (RUSSELL, 2019), é necessário encarar seu *design* ético como um direito fundamental. Afinal, os riscos que seus usos podem trazer são altos e incluem ameaças à democracia, à privacidade, criação ou reforço de preconceitos, dentre outros. Pois,

[...] como todo ente racional, o algoritmo aprende como modificar o estado de seu ambiente - nesse caso, a mente do usuário - com vistas a maximizar sua própria recompensa. As consequências incluem o ressurgimento do fascismo, a dissolução do contrato social que sustenta a democracia pelo mundo e potencialmente o fim da União Europeia e OTAN. Nada mal para uma linha de códigos, ainda mais com a ajuda de mãos de seres humanos (RUSSELL, 2019, p.9, tradução nossa).

²⁵ A expressão “massagear o modelo” foi abordada em palestra no 1º Congresso de Inteligência Artificial da PUC/SP. Disponível em: <https://www.youtube.com/watch?v=Xcm5TLvIyBg>. Acesso em: 12 de jan. 2021.

É essencial, portanto, atentar-se para a gestão de riscos e para a necessidade de apontar e reforçar caminhos éticos e o desenho de uma tecnologia em plena ascensão, como é a IA e cujas consequências e ameaças a direitos e liberdades individuais são amplas. Além disso, pressupor o *design* ético como direito fundamental objetiva assegurar que a IA seja humanocêntrica, isso é, que tenha o humano como centro e cujas regras éticas abranjam os atores envolvidos nos projetos e visem o desenvolvimento da cidadania dos indivíduos, bem como a construção de sistemas de IA confiáveis. Afinal, enquanto criadores e desenvolvedores da IA, somos nós, humanos, os responsáveis por uma IA responsável (DIGNUM, 2019).

O roteiro de desenvolvimento de uma IA ética que sirva para operacionalizar a auditabilidade dos programas implantados por instituições privadas e públicas precisa, entretanto, ser realista e de fácil execução, sob pena de, na complexidade, perder efetividade.

2 ÉTICA NA IA: OS *INPUTS* DA LITERATURA

O objetivo deste capítulo é entender e apresentar a discussão acadêmica atual sobre o tema da ética na IA por meio de análise bibliométrica inspirada em metodologia utilizada no artigo de Farranha e Bataglia (2021) e Soares et al. (2016). Por meio dessa observação, almejou-se identificar as direções, rumos e tendências da produção científica na área de ética e inteligência artificial. A interpretação dos dados permitiu verificar como temas envolvendo a ética estão sendo debatidos com maior frequência nas publicações internacionais.

Entender o debate científico pode contribuir para a criação da chamada Pesquisa em Inovação Responsável (PRI ou em inglês, RRI – *Responsible Research and Innovation*), apresentada na figura 5:

Figura 5– Processo de Pesquisa em Inovação Responsável



Fonte: (DIGNUM, 2019, p. 50, tradução nossa)

A chamada Pesquisa em Inovação Responsável fornece parâmetros para comunicação dos projetos de IA com a sociedade. Possibilita, desse modo, escrutínio público e críticas, abertura dos dados e movimentação a reflexão, adaptação, diversidade e transparência dos vários atores envolvidos na aplicação da IA.

Procurou-se então, neste tópico, mapear a produção acadêmica e literária sobre ética na IA em geral e a ética na IA aplicada ao Direito, por meio da realização de análise bibliométrica, pela qual, diante de um *corpus* de artigos e publicações científicas, buscaram-se métricas presentes nos textos e que apontam tendências acerca do tema (RIBEIRO et al., 2012).

Para seleção do *corpus* textual analisável, foram realizadas pesquisas no “Portal Capes” e selecionada a base de dados *Web of Science*, uma das Bases que na busca geral por assunto apresentou maiores números de publicações. Utilizaram-se como recorte temporal as publicações indexadas de 2017 a 2021 e selecionaram-se alguns descritores para a pesquisa.

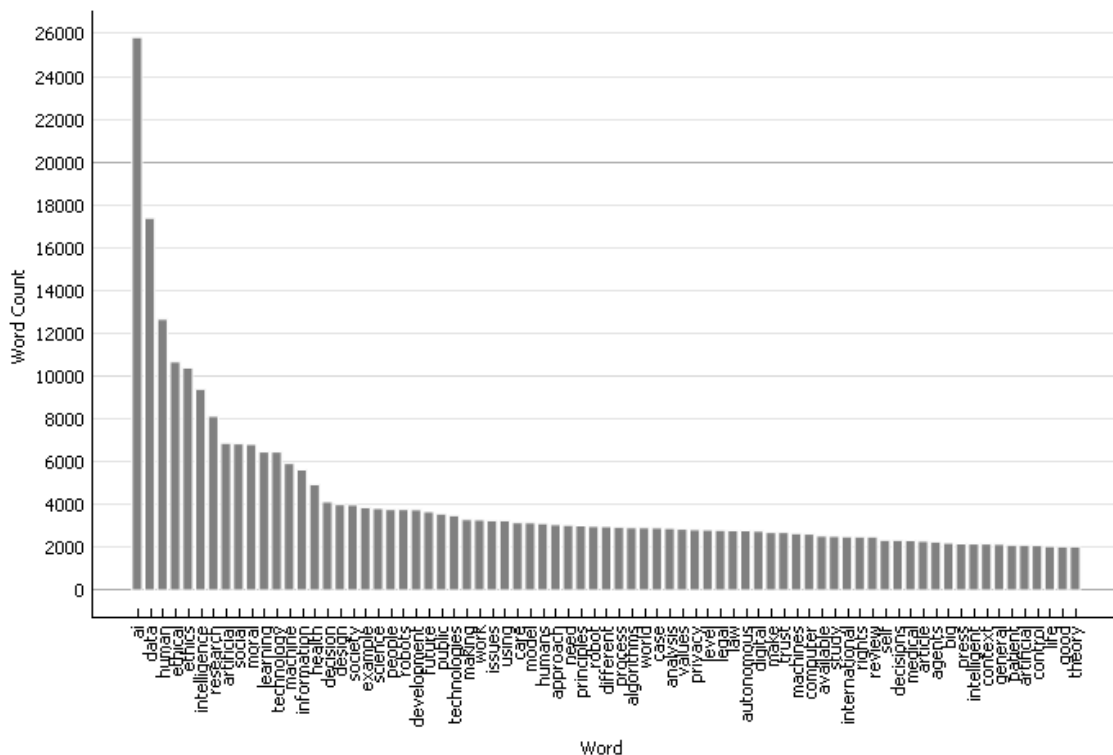
Os textos selecionados foram agrupados e salvos para, em seguida, serem carregados e explorados a partir do *software* “*Orange Data Mining*”, uma ferramenta que usa aprendizado de máquina para examinar determinado conjunto de informações. A interface apresenta diversas funcionalidades de IA para criação de fluxos, gráficos e demais atividades relacionadas à ciência de dados.

Na presente pesquisa, a ferramenta foi utilizada para identificar padrões textuais que levassem a obtenção de indicadores relativos à produção acadêmica sobre o tema ética e IA. Algumas questões que surgiram após a investigação e puderam ser contempladas foram: Quais os riscos éticos mais apontados pelos textos? Quais os princípios éticos trazidos pela literatura? Quais as áreas que mais aparecem? Quais as legislações mais recorrentes?

A resposta a essas questões apontou indicadores mencionados pela elaboração científica recente. Assim, identificaram-se temas e assuntos mais frequentes nas abordagens do tema da ética em IA.

Na primeira busca, foram selecionados os descritores “*ethics*” e “*artificial intelligence*”, com o intuito de encontrar aspectos presentes na literatura sobre ética para IA em geral. Foram encontrados 661 artigos e baixados 503 artigos cujo acesso estava disponível. Na Figura 6 observa-se a nuvem das 60 palavras que mais aparecem após o processamento dos textos completos localizados e carregados no *software Orange*, excluindo-se as expressões presentes no Anexo I.

Figura 6– Nuvem de palavras mais recorrentes nos textos completos a partir da pesquisa com descritores “*ethics*” and “*artificial intelligence*”

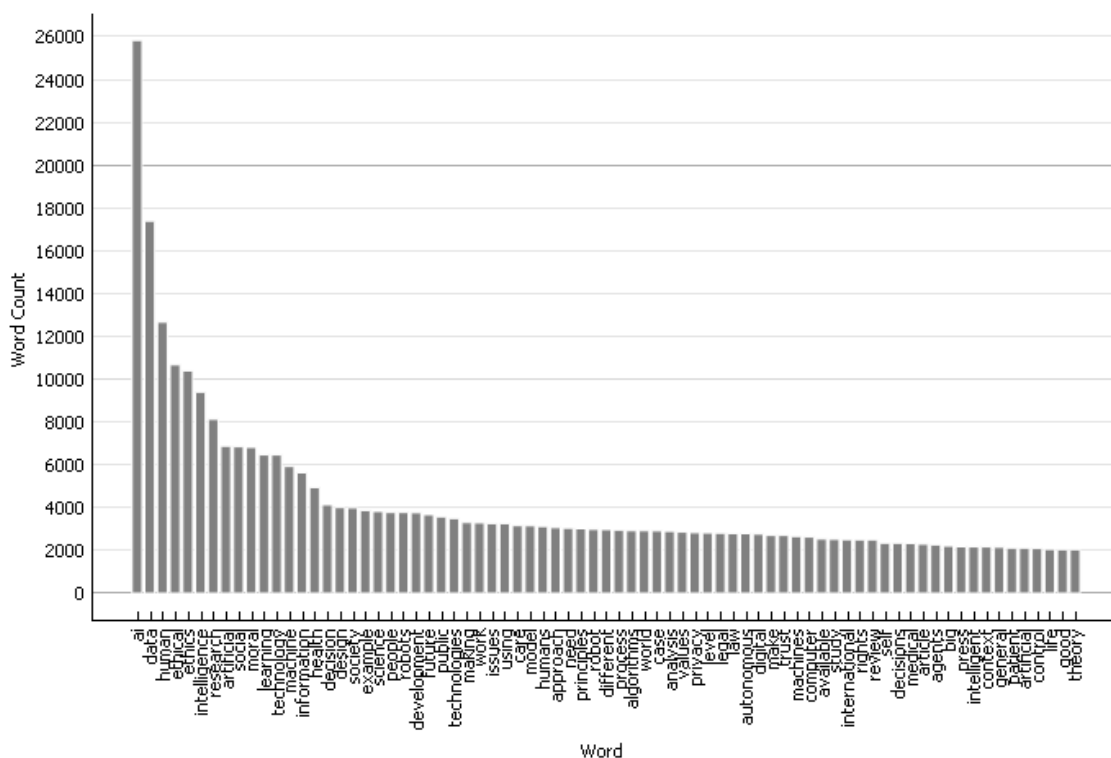


Fonte: elaboração própria

Como a pesquisa inicial abrangeu artigos que tratam de ética em IA, sem especificar a área do Direito, notou-se que entre as palavras mais frequentes aparecem “medicina”, “saúde”, “educação”, “segurança”, “direito”, “robôs”, “autônomos”, “decisão”, o que sinaliza os campos de conhecimento em que atualmente há maior preocupação sobre aspectos éticos da aplicação de IA da produção literária. Medicina e saúde são os campos que surgem com maior frequência nos textos analisados, o que aponta que a reflexão acadêmica tem se debruçado às questões éticas decorrentes das aplicações de IA nesse campo.

As palavras que descrevem problemas éticos da IA e que mais se repetem nos textos são “privacidade”, “dados”, “viés”, “sistemas”, “autônomos” e “responsabilidade”. A produção científica recente tem, portanto, se empenhado em debater como proteger a privacidade dos dados coletados para o desenvolvimento de IA, além de dedicar especial atenção aos vieses. Além disso, os sistemas autônomos e a responsabilidade desses é tema abordado de modo recorrente.

Os termos relacionados aos princípios que mais aparecem são “transparência”, “governança”, “responsabilidade”, “design”, “guias” e “princípios”. A maior discussão, como se percebe, é feita em torno da transparência, o que pode estar relacionado à preocupação com o uso de dados.



Fonte: elaboração própria.

Novamente a área que mais aparece nas palavras-chave é “medicina” e, dentro desse campo, os termos “radiologia”, “doença”, “câncer”, “paciente”, “bioética”, “covid”, indicando domínios na medicina em que a pesquisa em IA é mais desenvolvida e a preocupação com o cidadão paciente, sujeito à aplicação da IA na prática médica. Interessante notar também o debate que surge com a pandemia e o uso de IA no combate ao COVID-19, que aparece na produção bibliográfica recente.

Além desse assunto, também surgem nos resultados os termos “educação”, “direito”, “veículos” e “jornalismo”, que apontam outras áreas cuja pesquisa em IA é realizada e é mostrada nos artigos.

Os princípios e diretrizes mais recorrentes nos textos são “governança”, “política”, “transparência”, “prestação de contas”, “confiável”, “responsável”, “explicável”. Denota-se especial atenção à construção de uma IA compreensível, cujos resultados sejam explicáveis, inteligíveis e que permitam controle por aqueles a quem impacta.

Na segunda busca procurou-se selecionar descritores que envolvessem ética, IA e Direito, para obter informações presentes na literatura recente sobre a discussão da ética em IA aplicada ao campo do Direito. Nesse sentido, elegeram-se os descritores “*artificial intelligence*”, “*decision making*” e “*law*”. Foram encontrados 121 resultados

e baixados 73 artigos com acesso disponível. Os textos que resultaram da busca foram inseridos no *software Orange* que usa IA para criar métricas. Na figura 10 observa-se a nuvem de palavras que mais aparecem após o processamento dos 73 textos encontrados, excluindo-se as expressões presentes no Anexo II.

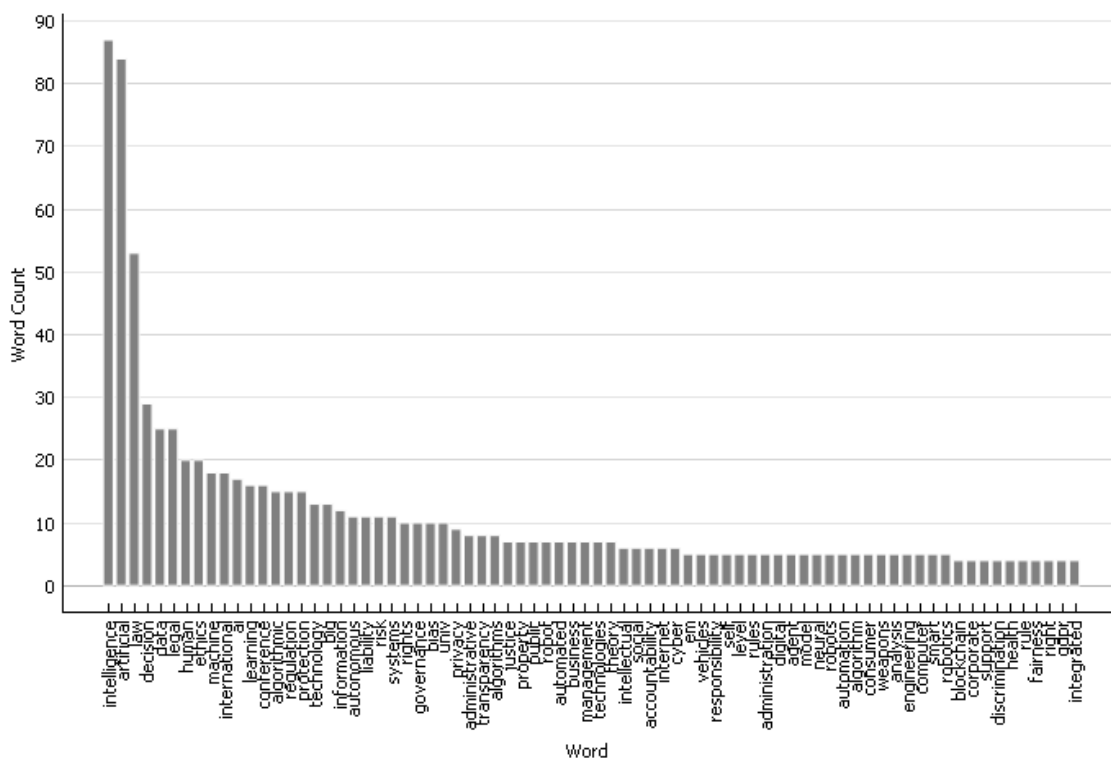
Figura 10- Nuvem de palavras mais recorrentes a partir dos textos completos da pesquisa com descritores “artificial intelligence”, “decision making” and “law”



Fonte: elaboração própria.

A figura 11 mostra gráfico elaborado com auxílio da ferramenta *Orange* a partir do número e nome das palavras que mais apareceram nos textos analisados.

Figura 11- Gráfico de palavras recorrentes nos textos completos a partir dos descritores “artificial intelligence”, “decision making” e “law”

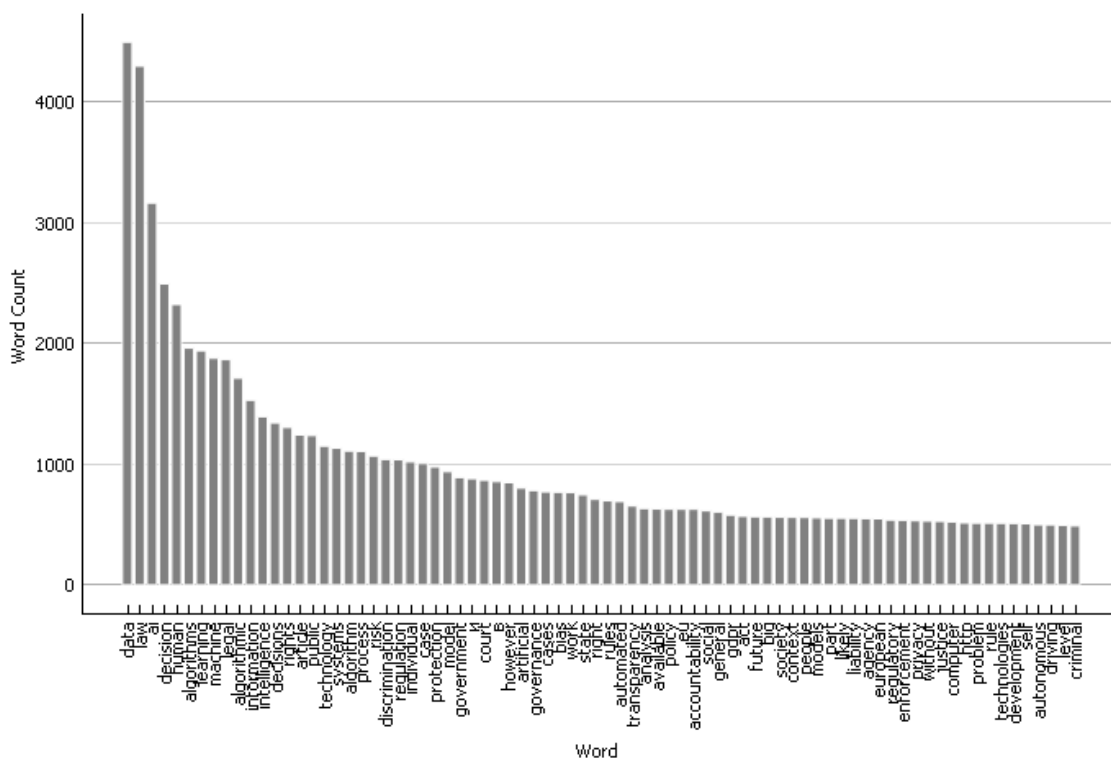


Fonte: elaboração própria

Nota-se que do resultado podem ser extraídas algumas variáveis como: quais os princípios éticos comuns nos textos? Quais os problemas trazidos pela aplicação de IA ao Direito aparecem comumente? Quais os tópicos do Direito mais citados? E as regras mais frequentes?

“Transparência”, “prestação de contas”, “equidade”, “responsabilidade” e “sustentabilidade” aparecem como princípios recorrentes na literatura analisada. A reiteração dos princípios que geram o acrônimo FAT (*fairness, accountability e transparency*, em português, equidade/não-discriminação, responsabilidade/prestação de contas, transparência e ética), demonstra que a academia prioriza esses três valores como norteadores da aplicação ética de IA.

Também são resultados importantes os termos presentes nos textos que identificam alguns problemas da aplicação de IA ou que demonstram pontos de inquietude trabalhados nos textos. Nesse sentido, são frequentes as palavras “viés”, “privacidade”, “dados”, “caixa-preta”, “preditiva”, “armas”, “julgamento”, “juiz”. A presença reiterada do tema dos dados e da privacidade, assim como a preocupação com a opacidade das decisões tomadas por IA (dificuldade de explicabilidade) nos julgamentos e previsões, indicam domínios em que são necessários maiores cuidados e atenção na elaboração de padrões e protocolos para aplicação ética de IA.

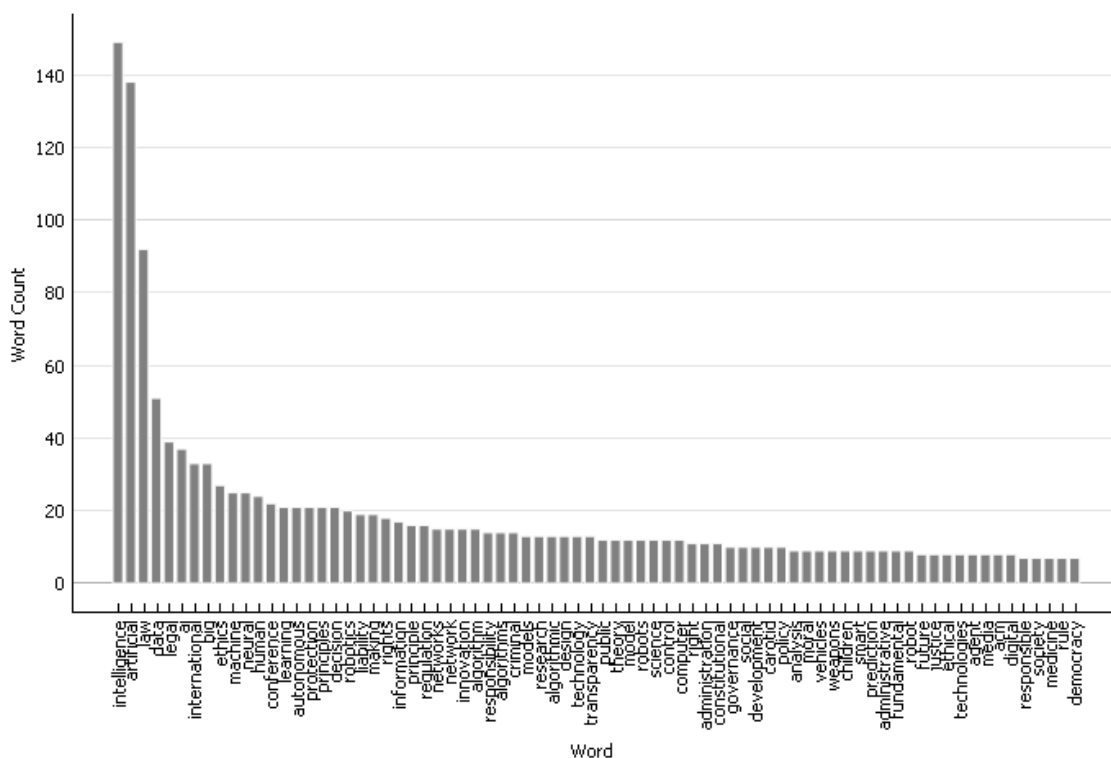


Fonte: elaboração própria.

Novamente destacam-se as expressões “transparência”, “prestação de contas”, “equidade”, “justiça”, como princípios mais trabalhados pela literatura. A reiteração das palavras indica a preocupação com a possibilidade e necessidade de controle da aplicação da IA, a qual pode trazer consequências preocupantes que se tornam abordagem comum também nos textos como “viés”, “caixa-preta”. Em especial, nessa pesquisa também aparecem os termos “criminal”, “autônomo”, “automatizado”, “preditiva”, “sentenciar”, “responsabilidade”. Manifestam-se, portanto, áreas de maior interesse da academia, as decisões automatizadas, principalmente no ramo do direito penal. E, outra vez, o “GDPR” aparece como regulamento bastante citado (580 ocorrências).

Por fim, a terceira busca utilizou os termos “*artificial intelligence*”, “*principle*” e “*Law*”. Encontrou-se 85 resultados e foram baixados 61 arquivos disponíveis. Ao importar os documentos para o *software Orange*, buscou-se averiguar nos textos completos quais as palavras mais recorrentes, que aparecem na Figura 14 (excluindo-se as expressões presentes no Anexo III).

Figura 14 - Nuvem de palavras mais recorrentes a partir dos textos completos da pesquisa com descritores “artificial intelligence” and “principle” and “Law”



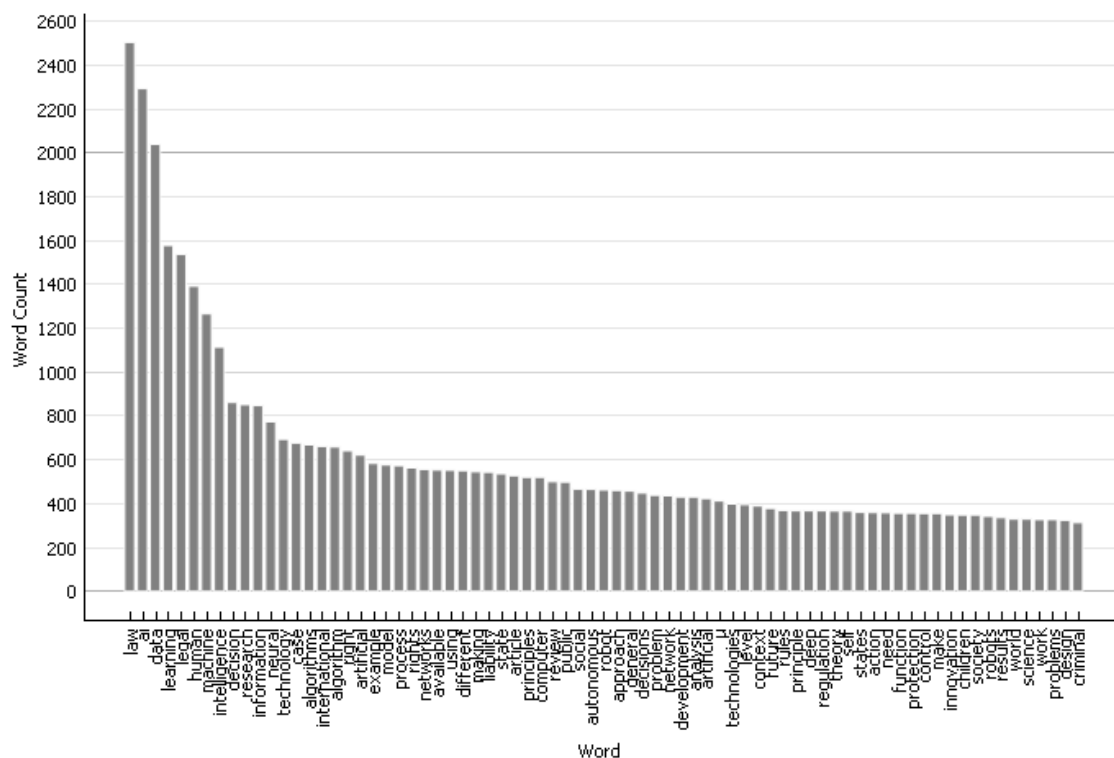
Fonte: elaboração própria.

O resultado apresenta pistas para compreensão de que as áreas do Direito que mais trazem inquietação na produção científica são “criminal”, “consumidor”, “administração”, “constitucional”, termos que apareceram de forma recorrente nos textos.

Denotam-se como expressões mais recorrentes, de igual forma, questões como “big data”, “dados”, “privacidade”, “proteção”, “julgamento”, “discriminação”, “viés”, “veículos autônomos”, “predição” e “caixa-preta”. Trata-se de consequências do uso da IA que inquietam a academia, eis que surgem em grande número nos debates promovidos e trazidos pelas publicações científicas.

A pesquisa aponta “transparência”, “governança”, “responsabilidade”, “controle”, “democracia” e “prestação de contas” como termos mais comuns relativos ao controle ético da aplicação de IA, assim como também registra 198 ocorrências de “GDPR”, regulamento europeu de proteção de dados.

Por fim, novamente utilizou-se o *software Orange* para a busca das palavras-chave mais frequentes e encontradas as expressões da figura 16.



Fonte: elaboração própria.

Evidencia-se, pelas palavras que aparecem em maior número, que os textos levantam a “acurácia”, “transparência”, “design” e “governança”, como norteadores principiológicos da aplicação de IA. O marco regulatório que mais aparece é mais uma vez o “GDPR”, que apesar de tratar sobre dados pessoais, demonstra ser mecanismo bastante debatido e considerado importante para quem discute o tema da IA.

Identificam-se pela análise a ocorrência das expressões “autonomia”, “armas”, “crianças”, “carros autônomos”, “tributos” e “criminal”, o que revela pontos em que a produção bibliográfica se debruça com maior atenção.

Nesse sentido, diante dos dados textuais minerados na ferramenta *Orange*, foi possível investigar como evoluiu o tema do controle ético na IA nos artigos publicados na base *Web of Science*, no período de 2017 a 2021.

Assim, tem-se que as produções científicas sobre IA têm sido maiores na medicina, que desponta como área do conhecimento com maiores ocorrências de artigos e trabalhos. A abordagem no campo do Direito centra-se principalmente na área criminal/penal.

Além disso, a evolução do tema na produção bibliográfica indica relevante preocupação com a privacidade, o uso de dados, os robôs e veículos autônomos, a

capacidade de predição, a discriminação, os vieses e as caixas-pretas que podem ser consequências da utilização de IA.

O constante aparecimento das palavras do acrônimo conhecido como FATE (*fairness, accountability, transparency e ethics*) (KEARNS; ROTH, 2020) revela também que o uso ético de IA deve contemplar a possibilidade de controle e governança das equipes contratadas para desenvolvimento, dos dados usados, do algoritmo criado e de todos os fatores capazes de conferir credibilidade do cidadão em relação ao uso da IA. E, ainda, o regulamento que mais aparece, apesar de tratar da regulação e proteção de dados pessoais e não especificamente de IA, é o GDPR, isso é, o respeito à intimidade e à vida privada é algo intimamente relacionado aos algoritmos de IA.

A partir da análise bibliométrica concluiu-se que os estudos sobre o controle da aplicação de IA abordam tópicos em comum e correlacionam-se entre si, na medida em que trazem como aporte para o debate temas similares. A reiteração de certos princípios e abordagens confirma que algumas noções são fundamentais quando se discute a boa execução dos projetos de IA na iniciativa pública e privada. Resultados estratégicos no uso dos algoritmos devem incorporar as demandas do *design* ético e precisam ser pensados tomando-se por base um mínimo de lições e *inputs* que a produção científica partilha e pretende nos informar.

3 INPUTS ÉTICOS NA INTELIGÊNCIA ARTIFICIAL: COMPARANDO DIRETRIZES

Este capítulo busca analisar e comparar as iniciativas regulatórias da CEPEJ e a Resolução n. 332 do CNJ para controle ético de IA aplicada ao Direito. Seu objetivo é extrair os indicadores dos sistemas de normatização²⁶, confrontando no que se complementam e no que são comuns ou no que divergem, além de apontar os desafios que essas normativas enfrentam.

Atualmente, existem algumas diretrizes tanto de organismos internacionais, quanto de alguns Estados nacionais e de empresas, que expressam a preocupação mundial com a questão do uso ético da IA de forma geral. A maioria baseia-se na *Conference on Beneficial AI* realizada pelo *Future of Life Institute* que criou os *Asilomar Principles* em 2017²⁷.

Dentre os organismos internacionais que se preocuparam em apresentar algum tipo de diretiva estão, por exemplo, o G-20, que elaborou princípios para IA Centrada nos Humanos, da Declaração Ministerial sobre Comércio e Economia Digital em 2019; a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que em 2019 lançou os Princípios da OCDE sobre Inteligência Artificial; as Diretrizes Éticas para uma IA Confiável de 2019, criadas pela Comissão de *Experts* de Alto Nível em IA da União Europeia (UE) e as Diretrizes da UE sobre ética em inteligência artificial de 2019. Além disso, a Unesco publicou em julho de 2020, recomendações sobre uso ético da IA²⁸.

Três anos após a publicação em 2019, de sua “Carta de Princípios”, a OCDE criou o *Framework for Classifying AI Systems* (uma estrutura para classificação dos sistemas de IA), que pretende ser uma ferramenta para ajudar legisladores, entes desenvolvedores e controladores a classificar os riscos e responsabilidades dos produtos da IA.

²⁶ A avaliação baseia-se em estudos como os de FEISTAUER et al., 2017 e VELOSO; MACIEL, 2015, que avaliam legislações na área de educação e no setor ambiental para extrair indicadores, parâmetros e critérios que guiam e norteiam essas áreas.

²⁷ O texto completo está disponível em: <https://futureoflife.org/2017/08/11/ai-principles/>. Acesso em: 08 jan. 2022.

²⁸ Em texto interessante, Santos (2020) expõe questões importantes sobre esse documento. Disponível em: . <<https://medium.com/deeplearningbrasil/highlights-from-unescos-recommendation-on-the-ethics-of-ai-7c641ab56ad5>>. Acesso em: 14 jul. 2020.

Trata-se de iniciativa fundamental que reconhece que cada projeto de desenvolvimento de IA tem suas complexidades próprias que precisam ser avaliadas. Além disso, serve de guia geral prática que possibilita análise concreta dos atributos econômicos envolvidos no projeto; da forma de coleta e utilização da base de dados; das pessoas e meio ambiente afetados; dos resultados esperados e obtidos e do modelo utilizado no desenvolvimento. Os critérios de avaliação criados são gerais, mas servem para compreensão dos riscos de cada projeto em concreto. Traduzem, portanto, o que pode-se considerar uma das iniciativas mais completas para controle ético da IA, posto que embora seja uma estrutura geral pode ser aplicada a todos os setores em que a IA é desenvolvida, como guia para produção de controle legislativo e de execução da análise, por órgãos fiscalizatórios.

Em 21 de abril de 2021 a UE divulgou outra iniciativa importante: uma proposta de regulamentação do desenvolvimento, implantação e uso da IA. A proposta é inovadora, pois pretende ser a primeira regulação que estabelece responsabilidades e sanções, não se tratando apenas de uma diretriz, e permanecerá em discussão por dois anos antes da aprovação final, para que o assunto esteja maduro. Por considerar a IA uma Tecnologia de Uso Geral (*General Purpose Technologies/GPT*) trata-se de documento bastante detalhado, que estabelece, por exemplo, quatro usos proibitivos de IA (para distorção e manipulação comportamental ou de grupos vulneráveis, para *social scoring* por autoridades públicas, para identificação biométrica remota em espaços públicos para fins de segurança²⁹). A Proposta traz a perspectiva classificatória da IA em três níveis de risco por uso risco inaceitável (*unacceptable risk*), risco elevado (*high risk*) e risco baixo/mínimo (*low/minimal risk*)³⁰. Os quatro usos inaceitáveis, por exemplo, são usos da IA classificados como de risco inaceitável (*unacceptable risk*).

Mais de quinze países também elaboraram documentos que de certa maneira traçaram balizas para o uso estratégico e responsável de IA e algumas organizações do setor privado que criaram autorregulações. Entre os países, tem-se, por exemplo, o Canadá, que em 2017 produziu a Declaração de Montreal para IA Responsável, além da

²⁹ Soares (2021) cita “três exceções que justificam a utilização destes sistemas, tais como: busca por vítimas específicas de crimes; prevenção a um risco específico, substancial e iminente como ataque terrorista, e detecção, localização, identificação ou investigação de suspeitos de ofensas criminais”. <https://www.jota.info/coberturas-especiais/inova-e-acao/proposta-de-actualizacao-regulatoria-para-ia-na-uniao-europeia-25052021>

³⁰ Em sua coluna na *Época Negócios*, Dora Kaufman critica a proposta, entendendo que a mesma ao invés de estimular a criação e desenvolvimento de IA traz rigidez à sua concretização. Disponível em: <https://epocanegocios.globo.com/colunas/IAgora/noticia/2021/04/proposta-europeia-de-regulamentacao-da-ia-impresoes-preliminares.html>. Acesso em: 14 dez. 2021.

Estratégia Pan-Canadense de IA em 2019. A Câmara dos Lordes, na Inglaterra, que propôs os “cinco princípios gerais para um código de IA”. A Alemanha lançou uma Estratégia para IA em 2018, por meio da qual traçou objetivos e planos para o bom desenvolvimento da IA ética.

Em 2019, em Pequim na China, uma união entre a Academia de Inteligência Artificial de Pequim (BAAI), a Universidade de Pequim, a Universidade de Tsinghua, o Instituto de Automação e Instituto de Tecnologia da Computação da Academia Chinesa de Ciências e empresas como Baidu, Alibaba e Tencent, criou o Plano de desenvolvimento de Inteligência Artificial da próxima geração. Nos Estados Unidos foram propostos princípios regulatórios de IA para uso no setor privado em 2020 (ITS, 2020)³¹. A *Google* lançou um “Guia Pessoas + IA”, escrito para ajudar profissionais a melhorar a experiência do usuário e realizar uma abordagem da IA centrada no humano³².

No contexto brasileiro deve-se ressaltar que, embora a Lei Geral de Proteção de Dados (LGPD) disponha sobre a salvaguarda do direito à privacidade e regule o tratamento dos dados pessoais, ela cumpre também papel importante e reflete no desenvolvimento ético da IA. Isso porque, como se comentou, a maioria dos modelos de IA lida com uma grande massa de dados, os quais em geral são necessários para sua criação e, desse modo, tangencia e esbarra em temas sensíveis³³ aos processos de automatização³⁴. Traz também a previsão imprecisa, no artigo 20, do direito “à revisão das decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade”, o qual tem íntima relação com o direito à explicação das decisões que utilizam IA (FRAZÃO, 2021a). Imprecisa porque não explícita como pode ser realizada e satisfeita essa revisão, nem tampouco especifica riscos das aplicações que merecem revisão.

Além disso, existem dois outros documentos importantes, um ainda em fase de concepção e outro recém-publicado no país. Um deles é a Estratégia Brasileira de

³¹ O Instituto de Tecnologia de Rio de Janeiro elaborou documento apontando diversas dessas diretrizes. Disponível em: <https://itsrio.org/wp-content/uploads/2020/03/RelatorioAI.pdf>. Acesso em: 23 out. 2021.

³² O Guia pode ser encontrado no site: <<https://pair.withgoogle.com/guidebook/>>. Acesso em 17 jul. 2020.

³³ Dados sensíveis são aqueles que merecem proteção especial, pois seu tratamento pode gerar perseguições, preconceitos e discriminações. São exemplos dados: de origem racial ou étnica, opiniões políticas, convicções religiosas ou filosóficas, filiação sindical, dados genéticos, dados biométricos, dados relativos à saúde ou dados relativos à vida sexual ou à orientação sexual.

³⁴ A Lei Geral de Proteção de Dados brasileira teve forte inspiração no Regulamento Geral de Proteção de Dados Europeu (RGPD), adotando, inclusive redação e princípios semelhantes (BIONI, LUCIANO, 2019).

Inteligência Artificial (EBIA), que recebeu de 12 de dezembro de 2019 a 31 de janeiro de 2020 percepções da sociedade por meio de consulta pública realizada pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC)³⁵.

A EBIA estabelece nove eixos temáticos, caracterizados como os pilares do documento. Três eixos horizontais: 1. Legislação, regulação e uso ético; 2. Governança de IA; 3. Aspectos internacionais, e seis eixos verticais: 4. Educação; 5. Força de trabalho e capacitação; 6. Pesquisa, desenvolvimento, inovação e empreendedorismo; 7. Aplicação nos setores produtivos; 8. Aplicação no poder público; 9. Segurança pública.

A EBIA chega tarde e por meio legislativo questionável (uma Portaria). Tenta demonstrar algum reconhecimento, pelo Estado brasileiro, da importância geopolítica da realização e proposição de políticas públicas de incentivo à concretização de ações estratégicas para desenvolvimento de IA genuinamente brasileira. Entretanto, tem recebido muitas críticas³⁶, pois apesar de ter ouvido diversos especialistas em audiências públicas e consultas (segundo Farranha (2020), foram 908 contribuições recebidas nas consultas), para assunto de tamanha relevância considera-se que as contribuições da mesma não foram efetivamente utilizadas para elaboração do documento. Além disso, faltam metas, programas e ações concretas que demonstrem o efetivo interesse brasileiro em efetivar uma estratégia de desenvolvimento tecnológico. Isso é, a EBIA não serve atualmente como documento de consulta geral e guia para demais setores que pretendem concretizar projetos em IA.

No Brasil, também tramita no Congresso Nacional o Projeto de Lei (PL) n. 21/2020, que estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências³⁷. O texto foi aprovado em regime de urgência na Câmara dos Deputados e agora está em tramitação no Senado.

Embora represente certo esforço regulatório por parte do Poder Público brasileiro, diversas críticas podem ser feitas ao Projeto, dentre elas o fato de que a matéria é complexa e exige amplas discussões e a despeito disso, foi aprovada na

³⁵ Disponível em: <<http://www.mctic.gov.br/mctic/opencms/inovacao/paginas/politicasDigitais/Inteligencia/Artificial.ht>>. Acesso em: 14 jul. 2020.

³⁶ Logo após a publicação do texto da EBIA, Ronaldo Lemos chama o documento de patético em sua coluna na Folha de São Paulo. Disponível em: <<https://www1.folha.uol.com.br/colunas/ronaldolemos/2021/04/estrategia-de-ia-brasileira-e-patetica.shtml>>. Acesso em: 02 mai. 2021.

³⁷ O texto completo do Projeto de Lei n. 21/2020 encontra-se disponível em: <https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928>. Acesso em: 14 jul. 2020.

Câmara em regime de urgência. Além disso, assim como a EBIA, não estabelece agenda de investimentos prioritários para o Brasil na área e tem previsões muito genéricas e pouco práticas. No Senado, por sua vez, foi constituída comissão de juristas para discussão e aprofundamento do tema e, espera-se, maior aprofundamento das discussões.

Se o intuito é criar um Marco Legal para aplicação da IA, o PL deveria servir como modelo de regulação geral e, portanto, não deixar dúvidas quanto à responsabilização dos agentes, tampouco quanto aos níveis de riscos de cada aplicação. Assim fazendo, poderia possibilitar que os demais setores desenvolvam as respectivas regulações específicas e sanções.

Apesar das críticas, como se pode perceber muitos documentos ao redor do mundo concebem balizas éticas para aplicação de IA e trazem a adoção de certos princípios. Pode-se identificar, na maioria dessas diretrizes regulatórias, quatro princípios principais sobre IA, que formam o acrônimo em inglês *FATE* (*fairness, accountability, transparency and ethics*), em português, equidade/não-discriminação, responsabilidade/prestação de contas, transparência e ética.

O relatório “Mapeamento de princípios da IA”, realizado por Burle e Cortiz (2020)³⁸, aponta seis principais dimensões principiológicas existentes: equidade (*Fairness*), Confiabilidade e Segurança (*Reliability & Safety*), Impacto Social (*Social Impact*), Responsabilidade (*Accountability*), Privacidade & Segurança (*Privacy & Security*) e Transparência (*Transparency*).

Um estudo realizado por Floridi e Cowls (2019), por meio do qual foram analisados nove códigos sobre boas práticas em IA, verificou a existência de mais de quarenta e sete princípios sobre o tema, o que, ao final, pode confundir mais que ajudar na imposição de limites éticos. Segundo a pesquisa, esses princípios podem ser concentrados e resumidos em cinco princípios fundamentais, cuja origem vem de princípios anteriormente propostos criados e aplicados na área da bioética, quais sejam: beneficência, não-maleficência, autonomia, justiça e o criado especialmente para a IA, explicabilidade.

Pode-se dizer que a beneficência pretende assegurar que a IA seja criada de forma sustentável e em benefício do Planeta, promovendo o bem-estar social. A não-maleficência consiste em evitar o mau uso da IA preservando-se direitos como a

³⁸ Disponível em: <<https://ceweb.br/publicacao/mapeamento-de-principios-de-inteligencia-artificial/>>. Acesso em: 14 jul. 2020.

privacidade e atendendo a uma série de precauções. O princípio da autonomia trata de garantir o processo de criação de IA humanizada de tal forma a impedir que o incremento de autonomia da IA signifique a perda de autonomia humana e possibilitando que o humano tenha o direito de decidir quem decide por ele (*the power to decide to delegate*) (FLORIDI; COWLS, 2019).

O princípio da equidade ou justiça compreende o cuidado com a promoção de diversidade e o resultado não-discriminatório no desenvolvimento dos projetos em IA. Por fim, a explicabilidade, princípio criado para a IA, também chamado de inteligibilidade das decisões, transparência ou *accountability*, é um complemento aos demais princípios, pois

[...] para que a IA seja benéfica e não maleficiente, devemos ser capazes de entender o bem ou o mal que está realmente causando à sociedade e de que maneiras; para que a IA promova e não restrinja a autonomia humana, nossa "decisão sobre quem deve decidir" deve ser informada pelo conhecimento de como a IA agiria em vez de nós; e para que a IA seja justa, devemos saber a quem responsabilizar no caso de um resultado sério e negativo, o que exigiria, por sua vez, um entendimento adequado do motivo pelo qual esse resultado ocorreu (FLORIDI; COWLS, 2019, p. 9, tradução nossa)

Buscar uma IA explicável é, então, potencializar o desenvolvimento dos demais princípios e perseguir o objetivo de, dentro das limitações de cada modelo, explicitar e informar os usuários sobre as possíveis falhas, a efetividade da aplicação, bem como os riscos que seu uso pode provocar³⁹. A explicabilidade é conceito vizinho e amplamente relacionado à noção de governabilidade/governança da IA e notoriamente busca o desenvolvimento de uma IA que seja confiável (LARSSON; HEINTZ, 2020).

A elaboração de mecanismos que busquem o controle ético da IA é tarefa complexa, pois exige que qualquer diretriz dialogue com a tecnologia. É fundamental compreender que a regulação que se pretenda eficaz deve contemplar os seguintes requisitos:

³⁹ O direito à explicação pode significar tanto informação sobre os riscos em jogo, quanto da razão da decisão tomada pelo algoritmo (modo como o algoritmo chegou a determinado resultado) (WACHTER; MITTELSTADT; FLORIDI, 2018). Entende-se ser mais eficaz e de mais fácil concretização o primeiro sentido. Nessa perspectiva, existem alguns níveis de implementação desse princípio: "(i) o *human-in-the-loop* (HITL), que se refere à capacidade de intervenção humana em cada ciclo de decisão do sistema, o que nem sempre é possível ou desejável, (ii) o *human-on-the-loop* (HOTL), que se refere à capacidade de intervenção humana durante o ciclo de design do sistema e o monitoramento da sua operação e o (iii) *human-in-command* (HIC), que se refere à capacidade de supervisionar a atividade global do sistema, incluindo seus impactos econômicos, sociais, jurídicos e éticos e a habilidade de decidir quando e como usar o sistema, o que pode incluir a decisão de não usá-lo em determinadas situações, de estabelecer níveis de discricionariedade humana durante o seu uso ou de assegurar a habilidade de contornar a decisão do sistema" (FRAZÃO, 2019, p. 514).

[...] Primeiro, deve usar fraca normatividade e não deve determinar universalmente o que é certo e o que é errado. Segundo, a ética da IA deve buscar proximidade com o objeto designado. Isso implica que a ética é entendida como um campo de estudo inter ou transdisciplinar, diretamente relacionado às ciências da computação ou às organizações da indústria adjacentes, e que é ativo dentro desses campos (DALY et al., 2019, p.7, tradução nossa).

Isso é, trata-se de entender que a legislação deve ser flexível o suficiente para se adequar e buscar permitir permeabilidade às constantes mudanças que são velozes. Além disso, deve ser elaborada por debates de especialistas multidisciplinares, de modo a incorporar diretrizes tecnologicamente viáveis e éticas, sob pena de serem ineficazes.

A dificuldade na regulação e governança da tecnologia é exposta pelo “dilema de Collingridge”: as consequências e impactos de uma tecnologia só podem ser medidas e perfeitamente conhecidas quando essa está em funcionamento e tem seu uso difundido; quando descobertos os efeitos negativos, no entanto, a regulação encontra entraves e resistência dos próprios usuários e empresas desenvolvedoras, tendo em vista que a tecnologia já está consolidada e os cidadãos já estão íntimos e de certa forma dependentes de sua aplicação cotidiana (COLLINGRIDGE, 1980).

Ebers (2019, p. 49, tradução nossa), aponta alguns instrumentos regulatórios importantes mencionados por diferentes autores que podem ser úteis na regulação de IA:

- Elaborar estatutos e diretrizes de maneira neutra em termos de tecnologia, a fim de garantir igualdade de tratamento e regras sustentáveis;
- Usar legislação multinível, especialmente combinando regras estatutárias com diretrizes que podem ser adotadas, avaliadas e alteradas facilmente pelos órgãos reguladores;
- Aumentar a flexibilidade por meio de “regulamentação temporária” usando “legislação experimental”;
- Criar zonas especiais para testes empíricos e desenvolvimento na forma de um laboratório vivo, ou "sandboxes regulatórios", nas quais o regulador fornece às empresas selecionadas que desejam comercializar produtos ou serviços inovadores com a oportunidade de implementá-los e testá-los em um domínio designado por um período especificado, sujeito a monitoramento e supervisão pelo regulador relevante, mas sem ser forçado a cumprir o conjunto de regras e regulamentos aplicáveis;
- Criar um “Comitê de Coordenação de Governança” para “supervisionar, cultivar o debate público e avaliar as ramificações éticas, legais, sociais e econômicas de (...) novas tecnologias importantes” ;
- Implementar “processos de feedback” em uma estrutura regulatória dinâmica que facilite o aprimoramento das informações para regulamentação, a fim de “permitir que os legisladores se adaptem às contingências regulatórias se e quando elas surgirem, porque um efeito de feedback fornece informações relevantes, oportunas, descentralizadas e informações específicas da instituição ex-ante;

- Aplicação de uma abordagem orientada a dados que permita a regulação dinâmica para identificar o que, quando e como regular. (Ebers, 2019, p. 49, tradução nossa)

A tarefa não é fácil, porém, é possível e envolve técnicas como a construção de meios para participação dos afetados pelos sistemas e de atores multiculturais na governança⁴⁰, estratégias como *sandbox* regulatórios⁴¹, para teste de tecnologias e elaboração de relatórios contínuos para avaliação de suas consequências e impactos dos projetos. Isso é, técnicas que protejam os direitos e liberdades fundamentais e, ao mesmo tempo, garantam e estimulem o desenvolvimento estratégico da IA (BAPTISTA; KELLER, 2016).

Em resumo, a aplicação de IA gera novos dilemas éticos e preocupações com a segurança de dados pessoais, o direito à privacidade, a transparência, a democracia, o preconceito e a discriminação, os quais impõem, de modo semelhante à bioética, a criação de novas maneiras eficazes de governança. Para que seja de fato considerada como AI4People ou AI4SG (*AI for People*, ou Inteligência Artificial para as pessoas e *AI for Social Goods*, ou Inteligência Artificial para o bem social), os avanços e benefícios trazidos pela IA precisam passar por regulações que a transformem em uma tecnologia eticamente apropriada (FLORIDI et al., 2020; FLORIDI et al., 2018).

3.1 GOVERNANÇA DA IA APLICADA AO DIREITO

A governança algorítmica é desafio que se impõe em todas as áreas do conhecimento em que a IA é utilizada e, como se viu, recentemente tanto a comunidade internacional como países e organizações privadas têm se mobilizado para criação de diretrizes regulatórias cujo objetivo é controlar e auditar os resultados e os meios de construção dessa tecnologia. A mobilização, entretanto, estão muito aquém da velocidade de criação dos avanços tecnológicos, o que tem levado a iniciativas realizadas sem o devido controle.

⁴⁰ A multiculturalidade na governança é essencial, segundo Dignum (2019), para garantia de que o controle e a transparência serão realizados de forma a se questionar por pontos de vista diversos as consequências éticas do resultado trazido pelo emprego do algoritmo.

⁴¹ Nome que remete às caixas de areia em que crianças podem brincar de forma segura. No caso da regulação seria um ambiente favorável ao teste seguro de melhores práticas de governança. O uso de Sandbox já tem sido aplicado para as novas tecnologias no Brasil, com previsão do Marco Legal das *Startups* (MICHILES, 2021).

Isso é, apesar da existência de diretrizes que podem ser chamadas de “universais” (em alusão às regras do direito internacional), que se destinam à governança algorítmica em geral, é interessante contar com normativas específicas para todo domínio em que a IA é executada. Afinal, em alguma medida, cada campo tem sua especificidade, seus dilemas e, por isso, suas questões sensíveis que merecem conformidade e controle. Do que tratam, então, as normativas que atualmente desenham a IA aplicada ao Poder Judiciário? Com o que se preocupam?

Para responder a essas questões, em primeiro lugar buscou-se compreender os âmbitos em que a IA tem sido empregada no Direito e, em especial, no Poder Judiciário. Em seguida, apresentaram-se quais e quantos projetos de IA vêm sendo desenvolvidos no Judiciário brasileiro e, por fim, foram analisadas a Resolução n. 332/2020 do CNJ e a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente de 2018, para entender o que essas diretrizes apontam como essencial no controle e governança da IA.

3.1.1 Usos da IA no Direito: concretização dos mandamentos do CPC/15

É certo que a IA não é a solução para todos os males da justiça, nem tampouco pode ser enxergada como algo completamente maléfico. Há, entretanto, um consenso a ser apontado: os algoritmos têm a capacidade de provocar intensas mudanças e transformações no mundo jurídico.

Isso porque, assim como os demais domínios da vida, o Direito não está alheio ao predomínio dos dados na sociedade da informação. Esses dados e metadados (extra) processuais de uma justiça cada vez mais digitalizada e eletrônica podem determinar os rumos da jurisprudência, da elaboração de instrumentos capazes de modificar a estrutura das profissões jurídicas e do modo como agem os personagens do processo.

No Brasil, existem importantes iniciativas estratégicas que demonstram que o Estado tem seguido uma tendência mundial⁴² e investido na digitalização de serviços públicos. Nesse sentido, em 29 de março de 2021, foi sancionada a Lei n. 14.129, por meio da qual se cria o chamado “Governo Digital”. A principal proposta do Governo Digital é a promoção da desburocratização por meio da prestação de serviço no

⁴² Países como a Estônia e a Itália têm se destacado na consecução de políticas públicas para implantação de Governos Digitais (Rossa, 2019).

ambiente digital, o que estimula a eficiência, bem como a criação de uma plataforma única de acesso capaz de facilitar o controle e transparência na gestão dos dados.

Acompanhando essa tendência do Poder Executivo, o CNJ tem capitaneado importantes iniciativas para a gestão de dados e promoção do acesso à justiça por meio digital. Desde 2013 a Resolução n. 185 instituiu o Sistema Processo Judicial Eletrônico - PJe como sistema informatizado de processo judicial no âmbito do Poder Judiciário e principal instrumento para tramitação dos processos. Nesse aspecto os profissionais da área jurídica viram-se desde já obrigados a adaptarem-se a atos processuais tecnológicos como assinatura digital, biometria, chaves privadas e públicas, documentos eletrônicos, intimações por vias eletrônicas etc. A paulatina mudança para o processo eletrônico foi iniciativa que encontrou já encontrou resistência desde o início de sua criação, até porque ainda não há uniformidade dos sistemas eletrônicos, o que pode dificultar o acesso à justiça e o exercício da advocacia.

Anos depois, por meio da Portaria n. 119/2019, foi criado o LIODS Laboratório de Inovação e Inteligência e Objetivos de Desenvolvimento Sustentável do CNJ, como espaço cooperativo de monitorar e desenvolver as questões da Agenda 2030 da ONU (Organização das Nações Unidas)⁴³.

Em 2020 veio a Resolução n. 335 de 29 de setembro que criou a Plataforma Digital do Poder Judiciário Brasileiro – PDPJ-Br e instituiu política pública para a governança e a gestão de processo judicial eletrônico, mantendo o sistema PJe como sistema de Processo Eletrônico prioritário do CNJ. A PDPJ-Br permitiu a unificação do trâmite de processos e garantiu a forma colaborativa entre os tribunais do país na gestão processual eletrônica⁴⁴.

As audiências por videoconferência tornaram-se ainda mais comuns a partir da pandemia mundial da Covid-19, as quais foram reguladas por resoluções como a n. 329 de 30 de setembro de 2020 (que estabeleceu critérios para a realização de audiências e outros atos processuais por videoconferência, em processos penais e de execução penal) e a Resolução n. 341 de 07 de outubro de 2020 que determinou aos tribunais brasileiros a disponibilização de salas para depoimentos em audiências por sistema de videoconferência.

⁴³ A Agenda 2030 da ONU buscou estabelecer Objetivos de Desenvolvimento Sustentável. <https://brasil.un.org/pt-br/sdgs>. Acesso em: 16 dez. 2021.

⁴⁴ Apesar de várias Resoluções e ações do CNJ para implantação do PJe, o desafio da estruturação de dados e saneamento dos processos no Brasil ainda é grande. Além de sistemas múltiplos de acesso ao PJe, que muitas vezes são de difícil manejo por advogados e partes, falta curadoria interna dos dados processuais (capaz de realizar uma boa classificação por temas das ações etc.).

Outro projeto importante foi introduzido em 09 de outubro de 2020 pela Resolução n. 345 do CNJ por meio da qual se criou o “Juízo 100% Digital”. A Resolução n. 345 veio acompanhada de cartilha que explica o conceito de “Juízo 100% Digital”:

[...] possibilidade de o cidadão valer-se da tecnologia para ter acesso à Justiça sem precisar comparecer fisicamente nos Fóruns, uma vez que, no “Juízo 100% Digital”, todos os atos processuais serão praticados exclusivamente por meio eletrônico e remoto, pela Internet. Isso vale, também, para as audiências e sessões de julgamento, que vão ocorrer exclusivamente por videoconferência (CNJ, 2020)

Outras Resoluções relativas à digitalização da justiça foram emitidas pelo CNJ em 2020 e coincidem com o período de gestão do Ministro Luiz Fux (2020-2022) que assumiu a presidência do CNJ com promessa de implantar ações para governança, eficiência, inovação tecnológica e transparência da justiça brasileira⁴⁵. Destacam-se aqui duas ações que representam a realização prática desse ensejo de modernização da justiça: a primeira foi a criação do “Programa Justiça 4.0 – Inovação e efetividade na realização da Justiça para todos”, cujo objetivo é “promover o acesso à Justiça, por meio de ações e projetos desenvolvidos para o uso colaborativo de produtos que empregam novas tecnologias e inteligência artificial”⁴⁶. A segunda, a Resolução n. 395, que instituiu a “Política Nacional da Inovação no âmbito do Poder Judiciário”.

Não se pode negar também que o contexto da pandemia do novo Coronavírus (Covid-19) acelerou a necessidade de rápida adequação do judiciário aos meios virtuais e não presenciais de garantia da prestação de serviços jurídicos que promovam o acesso à justiça. De todo modo, nota-se que o Brasil se esforça rumo à digitalização da justiça.

Garapon e Lassègue (2018) apontam que a Justiça digital é hábil a gerar verdadeira ruptura antropológica e revolução gráfica na atuação de juízes, advogados, réus, modificando a lógica do espaço-tempo da justiça. Isso porque, a partir das mudanças tecnológicas espera-se mais do Judiciário, vai-se da argumentação jurídica para a matemática (algorítmica ou baseada em estatística), altera-se a rotina de trabalho das figuras envolvidas no Direito e, por vezes, diminui-se ou exclui-se a necessidade do espaço físico. A justiça informatizada permite e serve de estímulo para o fortalecimento

⁴⁵ É possível ter acesso ao discurso completo no sítio eletrônico do STF. Disponível em: <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=451463>. Acesso em: 14 abr. 2021.

⁴⁶ O programa pode ser conferido no sítio do CNJ. Disponível em: <https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/>. Acesso em: 04 jun. 2021.

de políticas judiciais de implementação de IA para serviços na área do Direito e no Judiciário, na medida em que organiza e centraliza dados públicos.

Apesar de serem aplicáveis em variados contextos atualmente – de tradução de textos, ao reconhecimento facial – alguns pensam que a forma de raciocínio complexo do Direito impede que se usem técnicas de aprendizado de máquina ou aprendizagem profunda para a área jurídica. Há que se notar, entretanto, que a área jurídica é um campo fértil para consecução de modelos de IA, tendo em vista a grande massa de dados produzida nos conflitos pré e pós-processuais, os quais servem para treinar a aprendizagem de máquina e produzir resultados satisfatórios principalmente em atividades complementares à lida diária dos juristas, poupando tempo em tarefas repetitivas (SURDEN, 2014).

São muitas as funcionalidades possíveis de aplicabilidade da IA no Direito. Com base em revisão bibliográfica, foi possível criar um catálogo de opções desses usos transformadores. Pode-se dizer que tais usos identificados a seguir corroboram um conjunto de transformações desejadas desde algumas resoluções criadas pelo CNJ, apontadas anteriormente, mas principalmente e em especial à criação e aprovação do Código de Processo Civil de 2015 e seu conjunto de normas para uma justiça multiportas⁴⁷, eficiente, inovadora e atenta aos precedentes judiciais nas fundamentações judiciais.

Por meio de massa de dados, que podem ser obtidos por variados meios – como em casos processuais passados similares – treinam-se algoritmos para apontar padrões e prever, com base nesses indicadores e índices, probabilidades. O aprendizado de máquina supervisionado pode ser usado nesses casos para criar os chamados modelos preditivos ou *legal analytics*, que apontam probabilidades futuras de ação/decisão com base em dados passados. A também denominada QLP (*quantitative legal prediction* que no português pode ser traduzida como predição legal quantitativa - PLQ) utiliza larga escala de dados representativos e, assim, é capaz de medir performances de juízes e advogados; apontar valores de causa e expectativa de ganho/perda da ação (KATZ, 2012). Nesse sentido,

⁴⁷ Conforme explica Cunha (2021, p. 637) “a expressão multiportas decorre de uma metáfora: seria como se houvesse, no átrio do fórum, várias portas; a depender do problema apresentado, as partes seriam encaminhadas para a porta da mediação, ou da conciliação, ou da arbitragem, ou da própria justiça estatal”.

[...] esses sistemas são capazes de fazer uma variedade de previsões notáveis. No mundo do contencioso, quando há grandes conjuntos de documentos para revisar, fica claro que, em termos de precisão e recall, a revisão assistida por tecnologia (que tem sido classificada como uma forma de IA) pode agora superar os advogados juniores e extrajudiciais. Usando uma forma de "aprendizagem supervisionada", esses sistemas prevêem quais documentos advogados especialistas selecionariam como os mais relevantes. As previsões são baseadas nos documentos escolhidos previamente por especialistas de um conjunto de amostra. (SUSSKIND, 2019, p. 271, tradução nossa)

Uma das maiores expressões atuais dos modelos preditivos no Direito é a jurimetria, análise estatística que quando conta com a ajuda do aprendizado de máquina, pode ser útil para prever índices dos processos e inclinações decisórias dos juízes. A jurimetria tem como potencialidades a vantagem de oferecer previsibilidade ao julgamento das demandas e a promoção da “análise de risco da propositura (ou não) de uma demanda, de obtenção de um valor financeiro de acordo [...] sem olvidar da antecipação do resultado (ou não) de um recurso” (NUNES; DUARTE, 2020, p. 481).⁴⁸

Em determinadas situações a jurimetria pode gerar insights para exercício do que se chama de “fórum shopping” (previsto no art. 63 do Código de Processo Civil de 2015 (CPC/15)), ou escolha por contrato da jurisdição que as partes entendam ser mais vantajosa. Isso é, com base na análise dos dados processuais, aliados à potencialização causada pelo uso da IA, é possível compreender qual o foro julgará a causa em menos tempo ou atenderá melhor o interesse das partes, o que poderá gerar revolução no modo como se encaram os negócios (pré)processuais, ressaltando-se a necessidade de respeito ao acesso à justiça e paridade de armas processuais quando da consecução do negócio (art. 7º do CPC/15).

Surden (2014) aponta algumas limitações da jurimetria, como a necessidade de grande quantidade de massa de dados para melhorar a precisão do modelo, a obrigação de que se trate de casos semelhantes (ou seja, é inaplicável a casos difíceis ou com algumas nuances relevantes de diferença). Além disso há o perigo de hipergeneralização dos casos, ou seja, de que, no intuito de cumprir-se o mandamento do art. 926 do CPC/15 (manter a jurisprudência íntegra, estável e coerente), se aplique um precedente a casos com nuances de diferença, sem o devido respeito ao art. 489 do CPC/15 que determina a correta fundamentação da decisão judicial.

⁴⁸ A Associação Brasileira de Jurimetria (ABJ) é instituição sem fins lucrativos criada em 2011, composta por pesquisadores de diferentes áreas, que se tornou referência no país pela realização de estudos jurimétricos. A apresentação das análises empíricas tem potencialidades importantes como: apontar vieses e tendências de determinados juízes, ajudar no gerenciamento de riscos da ação, melhora a tomada de decisão e gestão estratégica da justiça e de litigância em escritórios de advocacia.

Predições feitas por aprendizado de máquina também têm sido usadas para reforço do sistema multiportas por meio de criação de soluções e plataformas para Resolução Online de Disputas (*ODR – Online Dispute Resolution*) (JUNQUILHO, 2020). “A comunidade de tribunais online têm se interessado mais em sistemas de AM que podem prever o resultado de decisões judiciais, muitas vezes com mais precisão, afirma-se, do que advogados humanos” (SUSSKIND, 2019, p. 272, tradução nossa). As ferramentas criadas podem dar subsídios argumentativos mais robustos para negociações (com base, por exemplo, nos possíveis resultados de eventual demanda judicial, isso é propor a "melhor alternativa para um acordo negociado" (*Best Alternative to a Negotiated Agreement - BATNA*) e usar técnicas compensatórias para facilitar a resolução do conflito (ZELEZNIKOW, 2017).

A prática diária do Direito lida ainda com um sem-número de documentos eletrônicos ou físicos como petições, contratos, intimações, guias etc. Por isso são tão vantajosas e comuns duas técnicas de aprendizagem de máquina que podem ser eficientes na organização desses documentos: a classificação automática e a *clusterização* (agrupamento).

A classificação automática permite encontrar documentos mais facilmente nos processos ou meios de provas, classificando-os como pertencentes a alguma categoria ou classe útil para as partes ou para servidores e juízes. Em determinados casos, “os algoritmos podem categorizar os documentos pela proximidade por meio da relevância (muitas vezes referida como "codificação preditiva" ou "revisão assistida por tecnologia)” (SURDEN, 2014, p. 113, tradução nossa). É possível, assim, por meio da análise de documentos, classificar, por exemplo, tipos de peças de um processo ou peças consideradas relevantes, irrelevantes ou provavelmente (ir)relevantes.

A abordagem da clusterização ou agrupamento no aprendizado de máquina também é útil para identificar relações entre documentos que possuem alguma característica semelhante em comum e associá-los, o que pode ser proveitoso para encontrar declarações, certificados, registros antigos e similares em um determinado conjunto de dados, por exemplo. “No *clustering*, um algoritmo de aprendizado de máquina tenta automaticamente agrupar itens que são semelhantes de alguma forma com base em alguma característica comum que o algoritmo detectou” (SURDEN, 2014, p. 114, tradução nossa). Nesse caso, é reforçado o mandamento fundamental da eficiência processual previsto no art. 8º do CPC/15, reduzindo-se o tempo de análise do processo por juízes, servidores e advogados.

O aprendizado de máquina também pode ser usado no Direito para desenvolver algoritmos capazes de realizar buscas de textos jurisprudenciais com maior precisão (MCGINNIS; PEARCE, 2014). Nesse sentido, o rastreamento de precedentes, além de ajudar a prática advocatícia, pode aumentar a confiança na argumentação apresentada pela parte na petição.

No âmbito da justiça brasileira, a precisão de buscadores jurisprudenciais também pode viabilizar a concretização dos mandamentos de integridade, estabilidade e coerência das decisões previstos nos arts. 489 e 926 do CPC/15. Isso porque apresenta àqueles que procuram situações semelhantes decididas no passado, o diálogo preciso e eficiente com questões similares anteriores. Em um sistema processual cada vez mais híbrido como o brasileiro, que traduz o conceito da antropofagia⁴⁹ do sistema *common law* ao tradicional *civil law*, a pretensão em assegurar maior segurança jurídica e estabilidade decisória mostra-se importante.

Algoritmos de aprendizagem de máquina também são capazes de gerar uma série de textos, rascunhos e fragmentos de modo automático. A automação de documentos é tecnologia que permite a produção em escala de arquivos como certidões, intimações, mandados, contratos em larga-escala, reduzindo tempo de elaboração e custos (SUSSKIND, 2017). Essa funcionalidade no Direito também pode assegurar que informações relevantes de cada caso não sejam desconsideradas por falha humana e otimizar o período de produções de peças processuais, principalmente em situações reiteradas e que se apresentam com certa constância no cotidiano judicial (reforçando também aqui a eficiência processual) (MCGINNIS; PEARCE, 2014, p. 45).

Considerando que Direito é linguagem e argumentação, quando se aborda a intersecção entre Direito e IA, não se pode deixar de compreender também o Processamento de Linguagem Natural (PLN ou *Natural Language Process* - NLP). “Processamento de linguagem natural (PNL) é o campo de criação de métodos e algoritmos que tomam como entrada ou produzem como saída dados de linguagem natural não estruturados” (GOLDBERG, 2017, p. 1, tradução nossa).

A linguagem natural traz inúmeros desafios para a linguagem da programação. É ambígua, traz *inputs* variáveis, símbolos com ligeiras discrepâncias, forma composicional, isso é, modificável pelo contexto e é esparsa, de forma que se altera ao

⁴⁹ (Junquillo & Jevaux, 2017) apresentam a necessidade de antropofagia dos precedentes ao sistema jurídico brasileiro, em especial a partir do CPC/15. O termo alude ao Manifesto Antropofágico de Oswald de Andrade.

unir-se com outros símbolos (outras letras, frases, palavras, sentenças etc.). O “Processamento de linguagem natural (PNL) é o campo de criação de métodos e algoritmos que tomam como entrada ou produzem como saída dados não estruturados de linguagem natural” (GOLDBERG, 2017, p. 1, tradução nossa).

Na área jurídica, da mesma forma, o sentido da linguagem expressa em textos, audiências e peças (sentenças, jurisprudência, acórdãos, ementas, atas de audiência, petições, entre outros), está em constante mutação (muda com o tempo, muda com o contexto, muda conforme cada uso⁵⁰) o que torna a aplicação de IA no Direito tarefa desafiadora. E qual a função do PNL? Essa metodologia é importante, pois

[...] permite que as máquinas se adaptem ao avaliar o texto [...]que o usuário identifique materiais que são provavelmente relevantes para sua pesquisa, mesmo se os materiais não contiverem palavras ou frases expressamente em sua lista de palavras-chave. O processamento de linguagem natural se aplica tanto à recuperação de informações (por exemplo, identificar a relevância de um documento) quanto à extração de informações (determinar os termos-chave de um documento). (ALARIE; NIBLETT; YOON, 2018, p. 8, tradução nossa).

Aqui não se pode deixar de mencionar o recém-descoberto algoritmo de PLN, *Generative Pre-training Transformer* (GPT), em sua terceira versão (GPT-03) lançada em 2020. O GPT-03 foi anunciado com 175 bilhões de parâmetros de dados para o aprendizado de máquina, o que traduz uma enorme potencialidade para desenvolvimento de textos. E em setembro de 2020 o jornal *The Guardian* publicou uma matéria⁵¹ totalmente escrita pelo GPT-03 que impressionou e assustou o mundo por sua contextualização e criatividade. O auspicioso algoritmo tem potencialidade no Direito para “simplificar rotinas, ampliar o acesso à justiça, eliminar filas de atendimento, e propiciar maior eficiência” (FALEIROS JR., 2021, p. 527).

Para finalizar a lista de usos da IA no Direito, pode-se dizer que uma das aplicações mais audaciosa seja a de produção decisória. A ousadia está no fato dessa ser uma das formas que geram maior temor de substituição do papel dos magistrados, que têm o papel de julgar os casos. A automação de decisões judiciais por algoritmos de aprendizado de máquina é feita com *inputs* de julgamentos anteriores, que trazem respostas com a previsão da solução jurisdicional (BORGESIU, 2020).

⁵⁰ Por exemplo: o sentido de denúncia no Direito Civil pode ser uma forma de encerrar o contrato e no Direito Penal pode ser noticiar um fato publicamente.

⁵¹ Disponível em: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>. Acesso em: 16 abr. 2021.

Essas têm sido as áreas em que se concentra a utilização moderna de AM no Direito, as quais foram resumidas no Quadro 1, a seguir.

Quadro 1– Lista de usos de aprendizado de máquina no Direito

Aprendizado de máquina em:	Usados no Direito para:	Artigos CPC/15
Modelos preditivos	<i>Legal Analytics</i> /Jurimetria, “fórum shopping”, ODR, identificação de fraudes, elaboração de alertas e criação de indicadores/métricas.	Arts. 63 e 926
Organização de documentos	Agrupamento, triagem e classificação de processos e peças; automação de rotinas internas aos tribunais e escritórios.	Art. 8º
Busca textual	Buscador de jurisprudências similares.	Art. 926 e seguintes
Automação de documentos	Geração automática de peças por transcrição de voz para texto ou com base em documentos anteriores escritos.	Art. 8º
Decisões automatizadas	Apoio à tomada de decisões; ou criação de sentenças, despachos e decisões jurídicas em geral.	Art. 489

Fonte: elaboração própria.

Como se nota, a transformação pelas novas tecnologias, em especial a IA, revoluciona, cria novos paradigmas e impõe desafios contemporâneos ou dá nova roupagem aos antigos dilemas ético-morais que permeiam o universo jurídico.

O uso da IA encontra solo fértil no Judiciário brasileiro abarrotado de dados processuais e, nesse sentido, os investimentos na área, reforçam a tentativa inicial do CPC/15 de desafogar o sistema jurídico. Isso porque, por meio do uso de uma “variedade de algoritmos e operando em grandes corpos de dados jurídicos, esses sistemas podem identificar padrões, regularidades e correlações que advogados humanos não conseguem quando usam métodos convencionais” (SUSSKIND, 2019, p. 271, tradução nossa).

Tanto é assim que o profissional esperado pelo mercado jurídico agora é chamado de 4 ou 5.0, aquele que está atento às novas tecnologias e que automatizam tarefas rotineiras e deixam para o humano o trabalho que desafia o intelecto e que

dependa de contato, comunicação, diálogo. O processo orientado a dados influencia diretamente, por fim, a produção do que vem sendo chamado de *visual law/legal design*, uso de uma linguagem mais amigável aos cidadãos nos diversos documentos jurídicos.

3.1.2. Questões éticas

Quem viu o filme *Matrix*, se lembra da cena em que Morpheus, personagem interpretado por Laurence Fishburne, oferece a Neo (Keanu Reeves) duas pílulas, uma azul, para regressar ao “mundo normal” e outra vermelha para permanecer no “mundo extraordinário”. Apesar das realidades distópicas apresentadas nos filmes de ficção científica, por muitas vezes apresentarem a máquina como algo diante do qual o humano pode perder o controle, a tecnologia não pode ser encarada como algo diabólico em si, de modo maniqueísta, mas como um instrumento que oferece importantes avanços e que deve ser balizado e submetido a limites éticos. Aqui não há obrigação de escolha entre uma pílula e outra, mas necessidade de seguir a fórmula ética em doses adequadas; isso é, nem a pílula azul nem a vermelha representam em si uma boa ou má escolha, o bom ou o mau uso da IA.

A variada lista de possibilidades do uso de IA na área jurídica mostra como seu uso pode trazer incontáveis benefícios. Cueva (2021) refere-se a essas mudanças como “promessas” da IA. Conforme o uso das novas tecnologias se desenvolve, se ganha em democratização do acesso à justiça e do próprio Direito, na medida em que os cidadãos passam a ter ao alcance maior gama de opções de serviços *online* com resultados mais rápidos e precisos. A harmonia no fluxo de recebimento e domínio da informação também permite a melhor *accountability* do serviço judicial. A mediação tende a se popularizar, cada vez mais oferecida por *legaltechs* e aumentar, assim, a possibilidade de soluções extrajudiciais e consensuais, além de proporcionar um reforço da previsibilidade das decisões e resultados, podendo gerar economia de tempo de serviços, permitindo a destinação de recursos a outros fins (CUEVA, 2021).

Ao afetar profundamente as profissões jurídicas, dentre as mudanças provocadas pela justiça digital, Garapon e Lassègue (2018) apontam três revoluções em andamento: a simbólica, a gráfica e a política. A revolução simbólica, o uso de dados pode trazer melhoria na produção das leis; a revolução gráfica, uma nova forma de escrita do Direito. Essa nova ordem inclui ainda uma revolução política que pode transformar a

máquina em “oráculo”, pois com seu uso espera-se mais do judiciário (*trust machine*) e confia-se no nível cognitivo as decisões preditivas (de bases matemáticas e não jurídicas).

Para a questão dos precedentes e a constante busca de incorporação (principalmente ao sistema jurídico brasileiro) (JUNQUILHO; JEVEAUX, 2017), o uso de IA traz muitas vantagens. Primeiro porque, ao padronizar determinadas decisões, permite que a atividade jurídica se concentre em questões complexas; segundo porque ao colocar parâmetros, estimula a manutenção da integridade, coerência e transparência decisória, aumentando a segurança jurídica e confiabilidade dos jurisdicionados e terceiro, porque pode estimular a comparação decisória entre magistrados e, assim, proporcionar a coerência jurisprudencial. Há também o fato de que torna a prestação jurisdicional mais rápida, aumentando a eficiência e, em alguns casos, reduzindo a imprescindibilidade de altos gastos com a existência ou manutenção do local físico e pessoal.

Em resumo, a IA, assim como os princípios no Direito, tende a ser ferramenta capaz de oferecer mudanças paradigmáticas na teoria e prática judicial. Seu uso, no entanto, precisa ser bem administrado, pois pode provocar alguns efeitos éticos adversos. Mas quais são essas implicações negativas e riscos trazidos pelo uso da IA, que exigem o equilíbrio e criticidade na concretização de projetos que envolvem o AM no Direito? Aqui novamente listam-se seis principais questões éticas cuja reflexão se impõe quando se aborda IA no Direito.

Inicia-se a avaliação pela discussão sobre consequências do uso dados pessoais na criação dos sistemas de IA. Em geral o processo de modelagem algorítmica envolve, como se sabe, a extração, armazenamento e análise de grande massa de dados que vão servir para aprendizagem da máquina e formam a *pipeline*⁵². Quando falamos de aplicações de IA ao Direito, são muitos os dados que podem servir como parâmetro para os usos apontados. A capa (página inicial) dos processos já traz em si uma abundância de informações que podem ser relevantes (metadados) como, por exemplo, o Cadastro de Pessoas Físicas (CPF), o nome das partes, o número do processo, o Tribunal, juiz ou órgão julgador, o Estado de origem do processo, o assunto ou tema. No interior do processo, encontram-se outras infinitudes de informações substanciais tais como a(s)

⁵² Termo usado aqui no sentido de caminho a ser percorrido pelo projeto de IA.

causa(s) de pedir, o(s) pedido(s), as peças processuais, o julgamento, decisões, valor da causa, gratuidade de justiça, pedidos de tutela provisória, dentre outros.

A primeira dificuldade gira em torno da obtenção desses dados. Embora públicos, a maioria dos (meta)dados processuais atualmente encontram-se desestruturados, desorganizados internamente em cada gabinete ou cartório, têm atribuída a classificação processual incorreta ou foram digitalizados e ainda não se submetem ao PJe (que no Brasil não é o único sistema de processo eletrônico).

Esse desalinhamento informacional dificulta a execução de ações no campo da ciência de dados, além de prejudicar o acesso à informação do jurisdicionado, o que representa, no fim das contas um entrave à ampliação do acesso à justiça (BECKER, 2019). Quanto aos processos digitalizados (que não nascem eletrônicos, mas precisam ser transportados para o meio virtual), há ainda mais uma dificuldade, pois demandam um procedimento chamado OCRização (OCR - *Optical Character Reader*) que realize a leitura e conversão da imagem de caracteres (Perez et al., 2017).

Isso é, faltam pessoal (equipes internas) e estrutura logística para fazer a curadoria interna organizacional dos dados e que cuide da publicidade precisa dos atos processuais, garantia constitucional prevista no artigo 5º, LX.

Outra controvérsia está em onde armazenar esses dados coletados. Aqui está presente a discussão de fenômeno denominado “colonialismo de dados” que, em resumo, traduz o monopólio de poderosas empresas e plataformas estrangeiras, na captura e armazenamento de dados (em especial os dados pessoais) de cidadãos de todo o mundo (CASSINO; SOUZA; SILVEIRA, 2021). Em maio de 2020, por exemplo, o TJSP rescindiu um contrato de R\$ 1,3 bilhões da *Microsoft*, após determinação de suspensão liminar pelo CNJ, sob o argumento de que a empresa licitante, sediada nos EUA, não poderia armazenar dados judiciais, reconhecidamente estratégicos para o Brasil⁵³.

Além disso, outra polêmica gira em torno de como gerir a aplicação da LGPD, para proteção de dados pessoais e privacidade da grande massa de dados processuais usadas no desenvolvimento de IA aplicada ao Direito.

Em primeiro lugar, episódios de vazamentos de dados e ataques cibernéticos ocorridos recentemente no Judiciário têm exposto a fragilidade dos sistemas de segurança e proteção dos dados da justiça brasileira, o que gera desconfiança da

⁵³ Disponível em: <https://www.conjur.com.br/2019-fev-21/cnj-suspende-contrato-tj-sao-paulo-microsoft>. Acesso em: 17 mai. 2021.

sociedade e do mercado em geral, quanto à eficácia da salvaguarda desse direito⁵⁴. Sendo a informação ativo estratégico de qualquer nação, investimentos em segurança da informação são questões básicas.

Em segundo lugar, é fundamental que sempre esteja presente cautela no desenvolvimento de IA quanto à necessidade de anonimização⁵⁵ de dados pessoais dos processos, muitos deles sensíveis (os quais têm proteção especial pelo art. 11 da LGPD, por representarem tipos de dados que podem levar a discriminações ou perseguições a grupos minoritários). Da mesma forma, deve haver especial preocupação com a preservação dos casos protegidos por segredo de justiça (previstos no sistema brasileiro, por exemplo, no art. 155 do CPC/15).

A concretização do amplo acesso à informação perpassa a discussão sobre a abertura da justiça (*Open Justice*), que se traduz na adoção de sistemas interoperáveis de processos eletrônicos, que facilitem a gestão e comunicação dos dados, por entender seu valor na sociedade informacional, inclusive como garantidor da transparência⁵⁶. É o que também defendem autores como Becker e Wolkart (2020)⁵⁷, que denunciam em seu texto a autodeclaração de fragilidade dos tribunais brasileiros, os quais se reconhecem sobrecarregados para consultas e uso massivo de dados. Os tribunais nesse sentido estabelecem entraves como o (*Re*)*Captcha*, para dificultar a captura das informações por empresas e não abarrotar os sistemas informacionais dos tribunais e, assim, acabam por dificultar o acesso à informação.

⁵⁴ Em 03 de novembro de 2020, o STJ noticiou um ataque cibernético de grandes proporções e em abril de 2021, o Tribunal de Justiça do Rio Grande do Sul (TJRS), sofreu ataques que suspendeu prazos e o funcionamento da justiça estadual. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/337701/39-dias-apos-o-ataque-cibernetico-ao-stj--reflexoes-e-desafios> <https://www.conjur.com.br/2021-abr-30/tj-rio-grande-sul-volta-alvo-ataque-hacker>. Acesso em: 08 mai. 2021.

⁵⁵ O art. 5º, XI da LGPD define o processo de anonimização como a “utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo”.

⁵⁶ É o que também defendem autores como Becker e Wolkart (2020), que apontam em seu texto a auto declaração de fragilidade dos tribunais brasileiros que se reconhecem sobrecarregados para consultas e uso massivo de dados. Os tribunais nesse sentido estabelecem entraves como o (*Re*)*Captcha*, para dificultar a captura das informações por empresas e não abarrotar os sistemas informacionais dos tribunais. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/regulacao-e-novas-tecnologias/como-roma-cidade-aberta-open-justice-21032020#_ftn3. Acesso em: 02 mai. 2021.

⁵⁷ Nunes e Malone (2021) também defendem a abertura de dados da justiça, que significam mais que a simples disponibilização de dados, mas que esses sejam apresentados de forma organizada, precisa e compreensível, eis que a mineração desses dados pode concretizar “tarefas de descrição – *description*: descrever os padrões e tendências reveladas pelos dados; classificação – *classification*: identificar a qual classe um determinado registro pertence; estimação – *estimation* ou regressão – *regression*: estimar o valor de uma determinada variável analisando os valores das demais; predição – *prediction*: descobrir o valor futuro de um determinado atributo; agrupamento – *clustering*: identificar e aproximar registros similares; e associação – *association*: identificar quais atributos estão relacionados (NUNES; MALONE, 2021, p. 137).

Ainda no que tange aos dados em especial usados para análise jurimétrica na área da justiça trabalhista, tem-se discutido em alguns julgados⁵⁸ a questão da paridade de armas no processo. Alegando “fraude por meio de algoritmos” e “dumping social”, alguns tribunais têm se recusado a homologar acordo e reconhecido existência de vínculo entre empregado e a empresa *Uber*⁵⁹. A jurimetria com o uso de IA pode representar uma gestão de dados mais eficiente, mas há que se ter, de fato, cautela para que essa não seja usada apenas por empresas que possuem poder aquisitivo para arcar com os custos de implantação ou desenvolvimento dessa tecnologia. A democratização do acesso à informação concretiza, portanto, paridade de armas processuais.

Quando se fala de IA aplicada ao Direito, outras duas questões éticas relacionadas, despontam: os vieses e a opacidade algorítmica. E o que são vieses?

Daniel Kahneman (2012), Nobel da Economia em 2002 e economista comportamental, é um dos autores que propõe em sua obra “Rápido e Devagar: duas formas de pensar”, que a mente humana funciona de dois modos diversos, os quais chama de Sistema 1 e 2. O mais comum é que as decisões humanas sejam tomadas pelo Sistema 1, forma rápida de pensar, que “opera automática e rapidamente, com pouco ou nenhum esforço e nenhuma percepção de controle voluntário” (KAHNEMAN, 2012, p. 29). Na forma de pensar do Sistema 1, o cérebro humano poupa esforços, tomando a decisão de forma precipitada, sem grandes reflexões. O Sistema 2, devagar, seria usado para corrigir erros do Sistema 1, demandando mais esforços intelectuais. O Sistema 2 deveria ser utilizado para solucionar questões mais complexas, “dominando os irrefreáveis impulsos e associações do Sistema 1” (KAHNEMAN, 2012, p. 29).

A forma de pensar rápida leva a conclusões precipitadas, tomadas por meio de atalhos cognitivos, chamados de “heurísticas”, isso é, diante de problemas “excessivamente complexos, e da impossibilidade de uma análise racional-compreensiva abrangente, indivíduos se valeriam de atalhos cognitivos, batizados como “heurísticas”, para solucionar problemas [...] que vieram a ficar célebres como “vieses”” (HORTA, 2019, p. 73).

⁵⁸ TRIBUNAL REGIONAL DO TRABALHO DA 4ª REGIÃO ACÓRDÃO 0020750-38.2020.5.04.0405 (PJe) RO; TRIBUNAL REGIONAL DO TRABALHO DA 3ª REGIÃO PROCESSO nº 0010258-59.2020.5.03.0002 (ED); TRIBUNAL REGIONAL DO TRABALHO DA 15ª REGIÃO 11ª CÂMARA ROT: 0011710-15.2019.5.15.0032. Disponível em: <https://www.tst.jus.br/>. Acesso em: 16 dez. 2021.

⁵⁹ “Isso porque, menos de 24 horas antes do julgamento, as partes juntaram petição requerendo homologação de acordo e, conseqüentemente, a retirada do processo da pauta de julgamento”. Disponível em: <https://www.conjur.com.br/2021-abr-28/desembargadores-trt-15-uber-manipula-jurisprudencia>. Acesso em: 16 dez. 2021.

O termo vieses surge, então, por meio da análise comportamental do pensamento humano e consiste em pré-concepções que levam o cérebro a concluir questões de forma irrefletida, sem basear-se em raciocínios lógicos. Também chamadas de heurísticas, os “vieses cognitivos”, atuam de maneira inconsciente de tal forma que “certas informações são privilegiadas, certas lacunas são inconscientemente completadas, certas percepções desencadeiam reações emocionais intensas, certas projeções de significado são rejeitadas pelos incômodos que nos provocam” (LINS E HORTA; COSTA, 2017, p. 18). Entender que o pensamento humano pode chegar a conclusões enviesadas significa compreender que podem atuar como fatores para a tomada de decisão de aspectos não racionais como os político-ideológicos, as identidades como gênero e raça ou os motivos contingentes como nojo e cansaço (LINS E HORTA, 2019).

Mas porque os vieses podem ser problemas éticos na criação algorítmica? Pois qualquer problema resolvido por criação algorítmica, em especial, o aprendizado de máquina, leva em consideração dados do passado. Assim,

[...] um algoritmo é tão bom quanto os dados com os quais trabalha. Os dados são frequentemente imperfeitos de uma maneira que permite que esses algoritmos herdem os preconceitos de tomadores de decisão anteriores. Em outros casos, os dados podem simplesmente refletir os preconceitos generalizados que persistem na sociedade em geral. Em outros ainda, a mineração de dados pode descobrir regularidades surpreendentemente úteis que são de fato apenas padrões preexistentes de exclusão e desigualdade. A confiança irrefletida na mineração de dados pode negar a participação plena de grupos historicamente desfavorecidos e vulneráveis na sociedade. Pior ainda, como a discriminação resultante é quase sempre uma propriedade emergente não intencional do uso do algoritmo, em vez de uma escolha consciente por seus programadores, pode ser anormalmente difícil identificar a fonte do problema ou explicá-lo a um tribunal (BAROCAS; SELBST, 2016, p. 671, tradução nossa)

Os vieses algorítmicos nascem, portanto dos dados passados utilizados no aprendizado de máquina não supervisionado e/ou das anotações dos humanos no aprendizado supervisionado, os quais podem conter e refletir algum tipo de preconceito que será reproduzido como padrão pelo resultado do modelo criado. Falar de vieses algorítmicos é assumir que humanos são preconceituosos, tendenciosos e tomam decisões de forma enviesada, o que se reflete na aprendizagem estatística⁶⁰.

⁶⁰ Em vídeo, Virginia Dignum, especialista em aplicação ética da Inteligência Artificial, explica de forma simples os vieses. Disponível em: < <https://www.youtube.com/watch?v=lgAUzCTdF7M>>. Acesso em: 18 jun. 2020.

No aprendizado de máquina não supervisionado os dados de entrada (*inputs*) não são previamente classificados por humanos, mas podem conter vieses e gerar ou reforçar resultados de agrupamento preconceituosos de todos os tipos (gênero, raça, religião, idade, dentre outros). Pelo aprendizado de máquina supervisionado, usado para predição ou tomada de decisões, o conjunto de dados pode ter sido classificado (intencionalmente ou não) de modo discriminatório pelo humano. Como, muitas vezes, não se consegue saber de antemão quais informações são ou não relevantes para certas decisões e quais podem gerar ou ampliar vieses algorítmicos, é mais fácil corrigir os *outputs* (saídas/resultados) (KEARNS; ROTH, 2020).

A inquietação com os vieses algorítmicos é tratada por Cathy O’Neil (2016), autora da área matemática que criou a sigla WMD’s (*Weapons of Math Destruction*) para denunciar uma das consequências do uso dos algoritmos, os quais denomina armas de destruição matemática. A criação de algoritmos, segundo a reflexão da autora, aumenta a desigualdade e ameaça a democracia, conclusão que Frazão também aponta em uma série de textos elaborados no ano de 2021 para uma coluna especial no Jota⁶¹.

Isso porque, os modelos de aprendizado de máquina se utilizam de dados do passado que podem conter vieses e, assim, refletir ou reforçar preconceitos contra aqueles que de antemão são grupos excluídos ou discriminados na sociedade como, por exemplo, pobres, negros e mulheres. Na perspectiva da autora, os vieses são perigosos, pois acabam por tornar os sistemas que se utilizam dos modelos de IA um círculo vicioso em que os privilegiados continuam sempre sendo privilegiados e os menos favorecidos ficam sempre excluídos (O’NEIL, 2016).

Esse sistema de repetição circular é chamado de efeito *feedback looping* e pode ser entendido como o fato de que a retroalimentação dos dados para criação algorítmica pode ampliar as diversas formas de discriminação, aumentando a exclusão social de minorias. Os modelos criados geram um espiral de discriminação que pode dificultar minorias nas mais diversas áreas da vida humana como na obtenção de crédito e empréstimos ou na obtenção de emprego (O’NEIL, 2016; KEARNS; ROTH, 2020).

Em geral, os modelos de aprendizagem profunda se utilizam de dados e produzem saídas (*outputs*) que são opacas e ininteligíveis. Além disso, muitas vezes não se consegue ter acesso ao porquê daquele modelo estar tomando determinada decisão (FRAZÃO, 2021).

⁶¹ Todos os artigos podem ser acessados no site da autora. Disponível em: <https://www.anafraza.com.br/noticias/>. Acesso em: 23 de out. 2021.

Para descrever esse fenômeno, Pasquale (2015) desenvolveu o conceito de *black box society* (sociedade da caixa preta), metáfora que trabalha com a noção de caixa preta de gravação e monitoramento ou que pode significar algo que “podemos observar suas entradas e saídas, mas não podemos dizer como um se torna o outro” (PASQUALE, 2015, p. 3, tradução nossa). Assim apresenta como problemas éticos no uso de dados para desenvolvimento de novas tecnologias, a relativização da privacidade e a dificuldade de explicação dos modelos algorítmicos.

A sociedade da caixa preta tem a vida observada e obtém decisões sobre diversas questões da vida por algoritmos, sem que os observadores forneçam informação sobre como essas decisões foram tomadas e quais dados pessoais foram usados. Verifica-se também uma distância cada vez menor entre Estado e mercado, os quais determinam a reputação dos indivíduos por algoritmos inescrutáveis e inacessíveis ou ininteligíveis (Pasquale, 2015).

Diante desses problemas, surgiram, inclusive, pesquisas que buscam a chamada IA explicável (da expressão em inglês *eXplainable AI* (XAI)). Com a finalidade de aumentar a confiança nos sistemas de IA e torná-los auditáveis a XAI

[...] propõe a criação de um conjunto de técnicas de ML que 1) produzem modelos mais explicáveis enquanto mantêm um alto nível de desempenho de aprendizado (por exemplo, precisão de previsão) e 2) permitem aos humanos entender, confiar adequadamente e gerenciar efetivamente a geração emergente de parceiros artificialmente inteligentes. XAI também extrai *insights* das Ciências Sociais e considera a psicologia da explicação. (BARREDO ARRIETA et al., 2020, p. 83, tradução nossa).

As pesquisas em IA explicável preocupam-se com o *design* ético e responsável da Inteligência Artificial (DIGNUM, 2019). O que se busca nesse campo é desenvolver o aprendizado de máquina com resultados cada vez mais precisos e que levem em consideração valores éticos como a equidade e transparência.

Nas aplicações de IA ao Direito, da mesma forma, a rotulagem e a escolha das variáveis ou conjunto de dados usados pode gerar resultados enviesados e opacos. Dentre os exemplos mais famosos está o caso de *State vs. Eric Loomis*, que envolve o uso do *software* COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizado em alguns Estados dos Estados Unidos para auxiliar os juízes por meio da aplicação de IA a avaliar o risco de reincidência dos réus em processos penais. Em 2013 *Eric Loomis* foi preso em flagrante no Estado de *Wisconsin* e julgado por um juiz que, baseado no resultado do *software* COMPAS, decidiu pela manutenção da

prisão do cidadão, negando a fiança. *Loomis*, então, recorreu à Suprema Corte de *Wisconsin* para questionar a decisão, argumentando que, em respeito ao princípio constitucional do devido processo legal, possuía o direito de acesso aos critérios que levaram o algoritmo a chegar à decisão pela negativa de fiança e alto risco do ofensor. A Suprema Corte de *Wisconsin*, entretanto, negou o direito de *Loomis*⁶², que levou a questão, por meio da utilização do *writ of certiorari* à Suprema Corte Americana. A Suprema Corte dos Estados Unidos considerou-se insuficientemente preparada para julgar a causa, devido aos aspectos tecnológicos e complexos que envolvem o tema e, assim, decidiu por não analisar a questão (FERRARI; BECKER, 2021).

Além do julgamento paradigmático sobre a plataforma, em 2016, a Organização Não Governamental (ONG) ProPublica também divulgou pesquisa em que denunciou a descoberta de que o uso do sistema COMPAS era discriminatório na avaliação de risco de reincidência, pois em relação a réus negros, por exemplo, o valor sugerido era maior⁶³.

O relatório produzido pela ONG tornou-se paradigmático e tem sido muito usado como exemplo quando se abordam vieses nos sistemas de IA utilizados na justiça, pois quebrou, de certo modo, o mito da imparcialidade judiciária e do julgamento justo/equânime, que historicamente compreendeu existir isonomia de tratamento dos cidadãos julgados pelas cortes norte-americanas.

O relatório também demonstra a complexidade envolvida na análise das ferramentas de IA que visam identificar eventuais vieses produzidos. Não é simples identificar vieses. Tanto é assim que a ProPublica precisou eleger para análise do Condado de Broward (Flórida), único local onde conseguiu obter autorização para a análise dos dados dos dados de dois anos (2013-2014) sobre a pontuação realizada pelo COMPAS.

Isso é, a avaliação mais conhecida de vieses nas aplicações de IA ao judiciário para apresentar conclusões relevantes e representar a quebra do mito da imparcialidade só foi possível porque a ONG obteve acesso aos dados usados e resultados obtidos; além de mobilizar tempo, recursos humanos e financeiros na pesquisa.

⁶² “A Suprema Corte de Wisconsin considerou que o uso de uma avaliação algorítmica de risco por um tribunal de primeira instância na sentença não violou os direitos do réu ao devido processo, embora a metodologia usada para produzir a avaliação não tenha sido divulgada ao tribunal nem ao réu.” Disponível em: < <https://harvardlawreview.org/2017/03/state-v-loomis/> >. Acesso em: 13 mai. 2021.

⁶³ O relatório com as conclusões e o modo de realização da pesquisa, está disponível em <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>>. Acesso em: 14 mai. 2021.

Quando se trata de uso do aprendizado de máquina em procedimentos judiciais, então, a discussão sobre vieses e caixas-pretas manifesta-se no direito à explicação e à fundamentação das decisões tomadas no contexto do Estado Democrático de Direito, no respeito ao devido processo legal, ao contraditório e à recorribilidade. A explicabilidade traz à tona o desafio regulatório em equilibrar as (im)possibilidades da tecnologia, às necessidades de limites éticos ao desenvolvimento de IA.

Mesmo quando aplicada a escritórios ou resoluções de conflito no âmbito privado, a IA deve ter por objetivo de alguma maneira melhorar acesso à justiça. Nesse sentido, há também preocupação com a transparência desses projetos. Quanto aos projetos desenvolvidos no setor público, a transparência envolve a comunicação informativa das propostas de IA, que propicie a governança dos mesmos pelos jurisdicionados afetados ou cidadãos que não estão diretamente envolvidos, mas que têm garantido o direito de acesso à informação (VALENTE; SUGAI; JUNQUILHO, 2021). A anatomia da transparência, expressão criada por Koivisto (2016), envolve interpretabilidade dos modelos, criação de explicações compreensíveis e acessíveis, isso é, o constante equilíbrio entre a linguagem técnica e a publicização adequada dos modelos (Larsson & Heintz, 2020).

No âmbito das questões processuais, há também a preocupação com o engessamento decisório e com uma possível parametrização exacerbada que produza decisões em massa e inadequadas às especificidades do caso concreto, no que vem sendo chamado por Nunes e Marques (2018) de “risco do decisionismo tecnológico”. Isto é, o algoritmo pode tomar certas situações como pertencentes ao caso geral, por vezes sem que a parte saiba por que foi inserida na generalização. Nesse caso, como a parte pode exercer plenamente a ampla defesa e o contraditório e auditar decisões comuns dadas para casos fora do padrão? Os julgamentos realizados com auxílio do aprendizado de máquina ferem de alguma maneira a independência e imparcialidade dos juízes?

Por fim, um dos maiores temores é o da substituição do trabalho humano, tanto de juízes como de advogados.

Quando esses sistemas não humanos de hoje são capazes de fazer previsões, identificar documentos relevantes, responder a perguntas e lidar com emoções em um padrão mais elevado do que os seres humanos, não é apenas razoável, é vital que perguntemos se, nas próximas décadas, pessoas ou sistemas estarão realizando uma ampla variedade de trabalhos que ocorrem em nossos tribunais hoje. (SUSSKIND, 2019, p. 274, tradução nossa)

E, nesse aspecto, não se pode negar que está em curso uma revolução no modo de trabalho do ser humano, a qual afeta todas as áreas do conhecimento, incluindo o Direito. Por outro lado, a lida direta com as soluções derivadas do aprendizado de máquina, como é o caso de algumas aplicações em ODR, pode tirar a confrontação pessoal e desumanizar, em certa medida, o acesso à justiça.

Em resumo, optou-se por listar os principais desafios éticos aqui analisados no Quadro 2 abaixo:

Quadro 2- Principais desafios colocados pela aplicação de IA ao Direito

Desafio geral	Consequências para o Direito⁶⁴
Uso de dados para aprendizado de máquina; Privacidade; proteção de dados pessoais	Dificuldade de acesso à informação processual; Disparidade de armas processuais; Falta de estruturação dos dados judiciais; Polêmica sobre uso de dados sensíveis; Onde armazenar dados coletados; <i>Open justice</i> e publicidade de atos processuais.
Vieses	Princípio da não discriminação; igualdade.
Opacidade/Caixas-pretas/Explicabilidade	Princípio da recorribilidade; ampla defesa e contraditório; fundamentação das decisões.
Transparência	Publicidade e controle da administração; acesso à informação.
“Decisionismo tecnológico”	Precedentes inadequados ao caso concreto
Perda de postos de trabalho e desumanização	Imparcialidade e independência judiciais; Substituição de advogados e juízes por funcionalidades mais precisas de IA.

Fonte: elaboração própria

Esses desafios colocam-se não como empecilho ao uso e criação tecnológica, mas como fatores, razões e justificativas principais pelos quais se torna fundamental refletir sobre a governança algorítmica. A concepção de limites éticos é necessária

⁶⁴ Um relatório sobre o uso de tecnologias durante a pandemia de COVID-19 alerta para o risco futuro como uso de *deep fakes* (criação, por técnicas de aprendizagem profunda, de rostos que reproduzem fielmente as características humanas) nas audiências virtuais, que podem afetar sensivelmente as provas testemunhais e depoimentos/declarações no processo. Esse risco não foi apontado aqui pois ainda é uma ameaça futura. Disponível em: <<https://bernardodeazevedo.com/wp-content/uploads/2020/08/final-white-paper.pdf>>. Acesso em: 17 mai. 2021.

quando se compreende que, como qualquer outra funcionalidade da justiça, a IA deve se submeter aos ditames do Estado Democrático de Direito.

3.1.3 IA NA JUSTIÇA BRASILEIRA

A primeira iniciativa de uso de IA no Direito foi a criação do robô ROSS. Concebido pelo sistema *Watson* da IBM (*International Business Machines Corporation*), ROSS utiliza PLN e computação cognitiva para apresentar a solução exata de pesquisa jurisprudencial realizada por perguntas feitas por advogados nos Estados Unidos. Além disso, monitora e notifica usuários de novas decisões judiciais sobre assuntos pesquisados. A tecnologia de pesquisa jurisprudencial e respostas jurídicas às buscas nasceu por demanda e iniciativa do escritório *Baker & Hostetler* de direito falimentar, nos EUA, mas atualmente têm recebido outros aportes e investimentos de tal forma que seu cofundador “Andrew Arruda, afirma que o ROSS está economizando até trinta por cento do tempo dos advogados” (BAKER, 2018, p. 15, tradução nossa). Desde então, nos EUA mais de 36% dos escritórios com mais de 50 advogados utilizam alguma tecnologia de IA (MARCHANT, 2017).

Na Europa, a Estônia anunciou em 2019 o início da criação e desenvolvimento de um robô juiz, capaz de julgar causas de até £7.000,00⁶⁵. A China também desponta como grande produtora de IA para o Direito, tanto é assim que criou sistema para revisão dos julgamentos por IA, além do chamado “mecanismo compulsório de busca e relatório de casos semelhantes”, o qual obriga juízes a fazerem buscas por jurisprudências semelhantes para manutenção da coerência e integridade do sistema judicial⁶⁶.

No Brasil, há esforços na área privada e no Poder Judiciário para implementação de modelos de IA que buscam concretizar funcionalidades aplicáveis ao Direito. Segundo pesquisa realizada por Magalhães (2021, p. 114), que pretendeu identificar a visão de cinco escritórios de advocacia brasileiros que utilizam IA em suas rotinas, “60% dos escritórios usam sistemas de IA para análise de documentos; 40% usam sistemas de IA para análise preditiva; e 40% dos escritórios usam sistemas de IA para

⁶⁵ Disponível em: <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>. Acesso em: 10 jun. 2021.

⁶⁶ Nesse texto, detalham-se as iniciativas Chinesas. Disponível em: <https://newlaw.com.br/na-china-o-futuro-do-poder-judiciario-esta-na-inteligencia-artificial/>. Acesso em: 10 jun. 2021.

atividades repetitivas”. A investigação concluiu que os grandes escritórios (com mais de 100 advogados) analisados, têm tentado incorporar rotinas de AM por meio da contratação de *startups* especializadas.

O desenvolvimento de tecnologias no Direito tem feito empresas especializadas em aplicação de soluções jurídicas nascerem para redução de custos, tempo e eficiência de rotinas jurídicas. Nesse sentido, a Associação Brasileira de *Lawtechs* e *Legaltechs* (AB2L) conta, em seu radar de associadas⁶⁷, com mais de cem *startups* cadastradas as quais utilizam *softwares* e desenvolvem respostas às dores e demandas do mercado do Direito.

“Estudo divulgado pelo “Comitê Gestor Nacional da Inovação do Poder Judiciário” mostra que 59 tribunais brasileiros têm pelo menos um laboratório de inovação na sua estrutura administrativa”⁶⁸. Além disso, existem atualmente, segundo aponta a primeira fase do “Relatório de Pesquisa: Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro”, realizado pelo Centro de Inovação, Administração e Pesquisa do Judiciário da Fundação Getúlio Vargas (FGV), 64 projetos de IA em aplicação ou em construção distribuídos em 47 tribunais do Brasil. No quadro abaixo, buscou-se sintetizar esses resultados⁶⁹:

Quadro 3– Análise das aplicações de IA no Poder Judiciário Brasileiro

TRIBUNAL	NOME DO PROJETO	EQUIPE DE DESENVOLVIMENTO	TECNOLOGIA DESENVOLVIDA
Supremo Tribunal Federal (STF)	Victor	UnB e STF	Victor: “[...] Ferramenta capaz de executar a identificação dos recursos que e enquadram em um dos 27 temas mais recorrentes de repercussão geral e a respectiva devolução aos tribunais de origem. Está habilitada para proceder à identificação e à separação das cinco principais peças dos autos: acórdão recorrido, o juízo de admissibilidade do recurso extraordinário,

⁶⁷ O radar dinâmico da AB2L separa aplicações por área e pode ser visitado em: <https://ab2l.org.br/radar-dinamico-lawtechs-e-legaltechs/>. Acesso em 17 mai. 2021.

⁶⁸ A reportagem é do Conselho Nacional de Justiça, que elogia a cultura da inovação nos tribunais brasileiros. Disponível em: <https://www.cnj.jus.br/laboratorios-de-inovacao-estao-presentes-em-59-tribunais-revela-estudo/>. Acesso em: 16 dez. 2021.

⁶⁹ Importante destacar que a pesquisa teve dificuldade em encontrar notícias sobre os projetos, o que demonstra que o Judiciário brasileiro apresenta problemas com a transparência dos mesmos.

			petição do recurso extraordinário, sentença e agravo no recurso.” (FGV, 2021)
Superior Tribunal de Justiça (STJ)	Athos; Sócrates; E-juris; TUA	STJ	<p>Athos: permite agrupamento automático por peças similares, a busca por peças similares, o monitoramento de grupos e a pesquisa textual. “Atua na rotina de identificação de acórdãos similares aos que já constam na base de dados de jurisprudência, a fim de que sejam agrupados, evitando-se, assim, a poluição da base. No Núcleo de Gerenciamento de Precedentes (NUGEP), a ferramenta de IA atua na identificação de processos que têm a mesma controvérsia jurídica, com vistas à fixação de teses vinculantes. O sistema também atua na identificação de matéria de notória relevância; entendimentos convergentes e/ou divergentes entre órgãos do STJ; possíveis distinções ou superações de precedentes qualificados”. (FGV, 2021).</p> <p>Sócrates: destinada aos gabinetes de Ministros também utiliza ferramentas do Athos.</p> <p>E-Juris: “realiza a extração das referências legislativas e jurisprudência citadas no acórdão do STJ para auxílio da tarefa de cadastro das que efetivamente embasaram os votos dos Ministros na composição do acórdão e de descarte das que foram meramente citadas; apontamento dos acórdãos publicados principais e sucessivos de mesmos temas jurídicos. É destinado à Secretaria de Jurisprudência.” (FGV, 2021).</p> <p>TUA: “Identificação do assunto do processo pelo sistema, de forma automática, para fins de</p>

			distribuição às seções do STJ conforme o ramo do direito em que atuam: Direito Público (Primeira), Direito Privado (Segunda) e Direito Penal (Terceira)” (FGV, 2021).
Tribunal Superior do Trabalho (TST)	Bem-te-vi	TST	Bem-te-vi: “Facilita a gestão de processos (classe processual, entrada nos gabinetes, avaliação das datas de interposição dos recursos) nos gabinetes dos Ministros” (FGV, 2021)
Conselho Nacional de Justiça (CNJ)	Plataforma Sinapses	TJRO e CNJ	Dentre as funcionalidades presentes na Plataforma, podemos elencar as seguintes: treinamento supervisionado para modelos de machine learning (classificação de documentos, extração de texto); versionamento de modelos, auditabilidade dos modelos; interface para importar datasets; ambiente multi-tenant; aprendizado por reforço. (FGV, 2021)
Tribunal Regional Federal da 1ª Região (TRF1)	Robô Secor; Banco de sentenças; Sistema de Inteligência de Busca (SIB); Análise Legal Inteligente (ALEI); Projeto execução célere	Secor: TRF1 e empresa Sonda Tecnologia Projeto execução célere: Acordo de Cooperação Técnica entre o TRF 1ª Região e o Conselho Federal da OAB Demais projetos: TR1	Secor: Levantamento de dados a serem enviados para o CNJ. Banco de sentenças: criação de banco de informações com buscas textuais precisas e céleres no conteúdo dos documentos judiciais SIB: ainda não utiliza IA. ALEI: aplicar métodos de aprendizado de máquina com o objetivo de usar seus potenciais do reconhecimento de padrões nos processos jurídicos relativos a julgamentos do TRF1. Projeto execução célere: plataforma de software de otimização da tramitação das demandas coletivas em fase de cumprimento de sentença ou execução (ainda em fase inicial). (FGV, 2021)

Tribunal Regional Federal da 2ª Região (TRF2)	Atendente virtual	TRF2	Atendente virtual: chatbot que simula uma conversa humana em um chat, no WhatsApp, e automatiza tarefas repetitivas, como dúvidas frequentes, na forma de diálogo pré-definido entre ele e o usuário.
Tribunal Regional Federal da 3ª Região (TRF3)	Sinara; Sigma (Sistema Inteligente de Gestão de Modelos de Admissibilidade);	TRF3	<p>Sinara: identifica textos jurídicos como leis, artigos, alíneas, e possibilita a pesquisa por assuntos, a fim de facilitar o trabalho nos gabinetes. (FGV, 2021)</p> <p>Sigma: produção de minutas de juízo de admissibilidade. Reconhecer a violação dos dispositivos constitucionais ou de lei federal pelo acórdão recorrido. Realizar a geração automática da produtividade de cada servidor.</p> <p>Prevenção: identificar possíveis casos de prevenção.</p>
Tribunal Regional Federal da 4ª Região (TRF4)	Classificação de temas na vice-presidência e turmas recursais; Análise de assunto dos processos; Triagem automática de processos a partir da petição inicial; Sugestão de modelos de minutas	TRF4	<p>Classificação de temas na vice-presidência e turmas recursais: Auxiliar o servidor na identificação do tema, apresentando uma ou mais sugestões de temas relacionados ao conteúdo do recurso para fins de análise do juízo de admissibilidade, pela Vice-Presidência e Presidência das Turmas Recursais, dos recursos destinados aos tribunais superiores em relação aos temas repetitivos do STJ, temas de repercussão geral do STF, pedidos de uniformização de jurisprudência da TNU, IRDRs e IACa do próprio tribunal, além de outros representativos de controvérsia.</p> <p>Análise de assunto dos processos: Classificação correta de assunto, evitando a necessidade de redistribuição por incompetência.</p>

			<p>Triagem automática de processos a partir da petição inicial: a partir da análise da petição inicial, identificam-se os casos de demandas repetitiva (projeto piloto em execução)</p> <p>Sugestão de modelos de minutas: criação de minutas, de forma a facilitar e padronizar a edição de documentos.</p>
Tribunal Regional Federal da 5ª Região (TRF5)	JULIA (Jurisprudência Laborada com Inteligência Artificial);	TRF5	JULIA: Auxilia na pesquisa jurisprudencial Uma segunda funcionalidade encontra-se em desenvolvimento, a fim de agilizar a identificação de processos sobrestados, cujas decisões devam ser reformadas em função de julgamento de processo judicial de um tema por Tribunal Superior (Repercussão Geral/Recurso repetitivo).
Tribunal de Justiça do Acre (TJAC)	LEIA (Legal Intelligent Advisor Precedentes)	Empresa Softplan e TJAC.	LEIA: identificar os processos vinculados aos temas de precedentes, para que magistrados e servidores validem ou não a sugestão de sobrestamento. Evitar a prolação de decisões diferentes para casos similares, vinculados a temas precedentes, bem como o dispêndio exasperado de tempo utilizado para a análise de processos. Sugerir, de forma automatizada, baseada na convergência entre o conteúdo da petição inicial de um processo e uma matriz de entendimento de um tema de precedente, a vinculação de processos a temas de precedentes.
Tribunal de Justiça de Alagoas (TJAL)	LEIA (Legal Intelligent Advisor Precedentes); Hércules.	LEIA: Empresa Softplan e TJAL. Hércules: TJAL e UFAL	LEIA: Vide TJAC Hércules: Agrupar petições semelhantes em uma fila específica, de forma que a criação de despachos ou de outros procedimentos necessários possa ser automatizada (projeto piloto em execução).

Tribunal de Justiça Amazonas (TJAM)	LEIA (Legal Intelligent Advisor Precedentes)	TJAM e empresa Softplan	LEIA: vide TJAC; Sugerir, de forma automatizada, para o advogado peticionante, o tipo da petição intermediária a ser juntada, de acordo com o seu inteiro teor; executar as ações de consulta, bloqueio e desbloqueio no Bacenjud, de forma automatizada e a partir da emissão de expediente no SAJ diretamente no site do Bacenjud, com o retorno das informações para o SAJ
Tribunal de Justiça da Bahia (TJBA)	Queixa cidadã	TJBA	Queixa cidadã: Realizar a identificação do requerente, comparando o documento de identificação utilizado na abertura da queixa com o usuário que está operando o aplicativo.
Tribunal de Justiça do Ceará (TJCE)	LEIA (Legal Intelligent Advisor Precedentes)	TJCE e empresa Softplan	LEIA: vide TJAC; Apoiar o trabalho intensivo em conhecimento de gabinetes para identificação de convergência entre as características dos processos judiciais em tramitação em suas unidades judiciais e as orientações de Tribunais Superiores em relação à vinculação aos temas (incluindo os respectivos <i>leading cases</i>)
Tribunal de Justiça do Distrito Federal (TJDFT)	Hórus; Ámon	TJDFT	Hórus: solução automatizada e inteligente, que auxilia nas atividades de identificação, classificação, correção, assinatura, carga e registro dos novos processos, que passarão a tramitar de modo digital. Ámon: identifica os visitantes na portaria do Tribunal a partir de fotos
Tribunal de Justiça do Espírito Santo (TJES)	3 projetos já em produção, ainda sem nome	TJES	Mineração de texto em base de dados não estruturada, com aprendizado de máquina em temas jurídicos relevantes. Plataforma inteligente de conciliação.
Tribunal de	IA332 (Identificação		IA332: Atender, de forma

Justiça de Goiás (TJGO)	automática 332) Sistema Automático de Identificação de Temas Repetitivos e Súmulas		<p>automatizada, os preceitos do artigo 332 do CPC.</p> <p>A solução busca identificar e classificar o processo, pela petição inicial no ato do peticionamento, sinalizando automaticamente se a presente ação judicial contraria:</p> <ul style="list-style-type: none"> • enunciado de súmula do Supremo Tribunal Federal ou do Superior Tribunal de Justiça; • acórdão proferido pelo Supremo Tribunal Federal ou pelo Superior Tribunal de Justiça em julgamento de recursos repetitivos; • entendimento firmado em incidente de resolução de demandas repetitivas ou de assunção de competência; • enunciado de súmula de tribunal de justiça sobre direito local. <p>Verificar pela petição inicial, no ato do peticionamento eletrônico, se a classe escolhida está coerente com a classe encontrada pela solução de IA.</p>
Tribunal de Justiça de Minas Gerais (TJMG)	Radar	TJMG	Radar: “identificar e agilizar os julgamentos dos casos repetitivos, em primeira e segunda instância” (ROSA; GUASQUE, p. 104)
Tribunal de Justiça do Mato Grosso do Sul (TJMS)	LEIA (Legal Intelligent Advisor Precedentes)	TJMS e UFGO	LEIA: Automatizar a identificação de possíveis casos de vinculação de temas de precedentes. Proporcionar isonomia no julgamento de processos similares e/ou de grande repercussão, redução do congestionamento do Judiciário e atendimento à Resolução 235 do CNJ.
Tribunal de Justiça do Mato Grosso (TJMT)	Ainda sem nome	Amazonas e TJMT	Robô assessor de magistrados, que elaborará minuta de sentenças.
Tribunal de Justiça do Paraná (TJPR)	Um projeto ainda sem nome; Larry	<p>Projeto sem nome: TJPR</p> <p>Robô Larry: Incubadora de Soluções Tecnológicas de Londrina e TJPR</p>	<p>Sem nome: Indicação de prescrição intercorrente de processos de executivo fiscal para todas as Unidades Judiciais do Estado do Paraná.</p> <p>Robô Larry: “auxiliar desde o trâmite até a sugestão de</p>

			sentenças em demandas de massa” (ROSA; GUASQUE, p. 116)
Tribunal de Justiça de Pernambuco (TJPE)	Elis	TJPE	Elis: ferramenta capaz de analisar e triar os processos de executivos fiscais, que totalizam mais de 50% de todas as ações que estão em trâmite em PE; criar um projeto de automação apoiado por IA para agilizar o processamento, bem como um <i>dashboard</i> para acompanhamento da evolução do processamento.
Tribunal de Justiça do Rio de Janeiro (TJRJ)	Ainda sem nome (FGV, 2021) Victoria (ROSA; GUASQUE, p. 98)	TJRJ	Ainda sem nome: em análise para implantação e criação Victoria: “automatizou os trâmites processuais das execuções fiscais” (ROSA; GUASQUE, p. 99)
Tribunal de Justiça do Rio Grande do Norte (TJRN)	Clara e Jerimum	TJRN e UFRN	Clara: “efetua a leitura de peças processuais e documentos e, a partir da interpretação que faz, ele sugere tarefas e recomenda decisões” (ROSA; GUASQUE, p. 102) Jerimum: “classificação e rotulação dos processos efetuada mediante análises com <i>deep learning</i> ” (ROSA; GUASQUE, p. 102)
Tribunal de Justiça de Rondônia (TJRO)	SINAPSE	TJRO	Automatizar tarefas repetitivas, inicialmente no Módulo Gabinete, por meio de ferramentas como predição do tipo de movimento processual, gerador de texto/autocomplete, identificação de seções em um acórdão e outras funcionalidades que agilizam o trabalho dos assessores e magistrados. Nos Juizados Especiais, será aplicada em casos repetitivos que representem um grande volume processual.
Tribunal de Justiça de Roraima	Scriba; Mandamus	Scriba: TJRR Mandamus: TJRR e UnB	Scriba: agilizar o processo de tramitação processual, especificamente na fase de

(TJRR)			instrução da audiência. Mandamus: o sistema usa a IA em três etapas: <ul style="list-style-type: none"> • análise da decisão • confecção do mandado • distribuição dos mandados que classifica por urgência, natureza, complexidade e geolocalização dos endereços.
Tribunal de Justiça do Rio Grande do Sul (TJRS)	Ainda sem nome	TJRS	Automatizar a análise inicial da demanda nos processos de executivos fiscais, a partir de técnicas de mineração de dados, associadas à tarefa de classificação, que permitam prever qual tipo de despacho deverá ser proferido nesta etapa inicial do processo judicial.
Tribunal de Justiça de Santa Catarina (TJSC)	Ainda sem nome	TJSC	Ainda em etapa de ensaio.
Tribunal de Justiça de São Paulo (TJSP)	Judi; LEIA (Legal Intelligent Advisor Precedentes)	Judi: TJSP e <i>Microsoft</i> LEIA: TJSP e Softplan	Judi: chatbot que conta com aprendizado de máquina. Nasceu da constatação de que, no atual contexto de relações contratuais e de consumo mais complexas e dinâmicas, os cidadãos têm cada vez mais dúvidas sobre as situações que podem ser solucionadas por meio dos Juizados Especiais Cíveis, levando ao esgotamento dos canais de atendimento dessas unidades judiciais – pois são, obrigatoriamente, presenciais. LEIA: vide TJAC
Tribunal de Justiça de Tocantins (TJTO)	Minerjus	TJTO e UFT	Minerjus: Classifica as petições iniciais de acordo com a TPU.
Tribunal Regional do Trabalho da 1ª Região (TRT1)	Inteligência Artificial e eficiência do judiciário;	TRFT1	Utilizar aprendizagem profunda (um subtipo avançado de inteligência artificial) na criação de modelos computacionais capazes de efetuar três tipos de análise preditiva: <ol style="list-style-type: none"> a) probabilidade de sucesso em audiência de conciliação; b) probabilidade de reversão ou modificação das sentenças proferidas pelas

			varas do trabalho; c) probabilidade de reversão ou modificação dos acórdãos proferidos pelas turmas do TRT/RJ. O modelo preditivo será implementado por meio de uma application programming interface (API), que poderá ser facilmente incorporada ao sistema PJe ou a qualquer outro de interesse do Tribunal.
Tribunal Regional do Trabalho da 4ª Região (TRT4)	Clusterização de processos	TRT4	Agrupamento de processos semelhantes com o propósito de agilizar a análise de admissibilidade dos Recursos de Revista. Agrupamento de processos semelhantes nos gabinetes dos Desembargadores a fim de otimizar a elaboração de voto.
Tribunal Regional do Trabalho da 5ª Região (TRT5) TRT 7ª Região, TRT 15ª Região e TRT 20ª Região	Gemini	TRT 7ª Região, TRT 5ª Região, TRT 15ª Região e TRT 20ª Região	Reduzir o esforço humano e o tempo gasto para identificação e agrupamento dos recursos ordinários similares, pendentes para julgamento, pois atualmente essa tarefa é realizada pelos servidores do Gabinete, mediante a leitura de recurso por recurso.
Tribunal Regional do Trabalho da 11ª Região (TRT11)	B.I. TRT 11	TRT 7ª Região, TRT 5ª Região, TRT 15ª Região e TRT 20ª Região	Análise de dados para comparações e insights de performance do Tribunal, auxiliando também nas correções e no monitoramento de atividades processuais do Tribunal. Predição de performance baseada em informações do passado. Indicação de itens de melhoria (alerta) baseados em performance do passado, de forma a atingir metas previamente configuradas em determinado prazo.
Tribunal Regional do Trabalho da 12ª Região (TRT12)	CONCILIA JT	TRT 7ª Região, TRT 5ª Região, TRT 15ª Região e TRT 20ª Região	Redução do tempo médio de duração de um processo na fase de conhecimento. Reconhecimento de processos com potencial para conciliação. Otimização de pauta.

Fonte: elaboração própria a partir do “Relatório de Pesquisa: Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro”, realizado pelo Centro de Inovação, Administração e Pesquisa

do Judiciário da Fundação Getúlio Vargas (FGV). Disponível em: https://ciapi.fgv.br/sites/ciapi.fgv.br/files/estudos_e_pesquisas_ia_lafase.pdf. Acesso em: 17 mai. 2021. E também em ROSA, Alexandre Moraes da; GUASQUE, Bárbara. O avanço da disrupção nos tribunais brasileiros. In: Inteligência Artificial e Direito Processual: Os Impactos da Virada Tecnológica no Direito Processual. Coord. Dierle Nunes, Paulo Henrique dos Santos Lucon, Erick Navarro Wolkart. Salvador: Juspodvim, 2021, p. 93-121.

A partir do ano de 2018 (FGV, 2021), portanto, nota-se um esforço do Judiciário brasileiro de adotar, em várias instâncias, projetos que utilizem modelos de IA envolvendo, na maioria dos casos, as próprias equipes de tecnologia dos tribunais. Em alguns projetos, há realização de parcerias com Universidades Federais e empresas privadas. Além disso,

[...] De forma geral, os projetos de IA nos tribunais comportaram as seguintes funcionalidades: verificação das hipóteses de improcedência liminar do pedido nos moldes enumerados nos incisos do artigo 332 do Código de Processo Civil; sugestão de minuta; agrupamento por similaridade; realização do juízo de admissibilidade dos recursos; classificação dos processos por assunto; tratamento de demandas de massa; penhora on-line; extração de dados de acórdãos; reconhecimento facial; chatbot; cálculo de probabilidade de reversão de decisões; classificação de petições; indicação de prescrição; padronização de documentos; transcrição de audiências; distribuição automatizada; e classificação de sentenças. Os principais objetivos atendidos pela inteligência artificial são: otimização de atendimentos aos advogados e ao público; maior segurança; automação de atividades; melhor gestão dos recursos humanos para a atividade-fim do Judiciário; aumento da celeridade na tramitação processual (FGV, 2021, p. 69).

Em resumo, os projetos de IA no Judiciário brasileiro buscam agilizar tarefas repetitivas, uniformizar o tratamento de peças e decisões processuais, estruturar dados, incrementar o sistema de precedentes e a eficiência da entrega da prestação jurisdicional.

3.1.4 O que trazem as diretrizes Europeia e Brasileira?

As diretrizes gerais que buscam impor balizas éticas à aplicação de IA ao redor do mundo⁷⁰ são genéricas, pois buscam limitar o uso de uma “Tecnologia de Propósito Geral (TGP)” (ARBIX, 2020). Isso é, de uma tecnologia cujo desenvolvimento se dá por meio de pesquisas que geram resultados impactantes em variadas áreas do conhecimento humano e transformam a sociedade em seus diversos aspectos, afetando a economia, a cultura, as relações humanas etc.

⁷⁰ Ver tópico 3.

No Direito, como apontamos nos tópicos 3.1.1 e 3.1.2 deste capítulo, existem igualmente diversos modos de utilização da IA. Esse uso, no entanto, pode trazer consequências éticas indesejáveis para o sistema de justiça e que não condizem com os ditames do Estado Democrático de Direito, o qual exige respeito a direitos fundamentais como a igualdade, o julgamento justo e imparcial, a não-discriminação, a fundamentação das decisões judiciais, a ampla defesa e o contraditório etc. Em outros termos, o uso de técnicas de AM no Direito pode gerar problemas singulares para a justiça e exigem uma regulação setorial, adequada, capaz de compreender as especificidades dessa aplicação.

Nesse sentido, o objetivo deste tópico é analisar as iniciativas regulatórias da Comissão Europeia para a Eficácia da Justiça (CEPEJ) e a Resolução n. 332 do CNJ para controle ético de IA aplicada ao Direito.

Inicia-se a análise pela Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais (CEEUIA SJ) e seu ambiente, adotada pela CEPEJ na sua 31.^a reunião plenária (Estrasburgo, 3 e 4 de dezembro de 2018), primeiro documento de controle específico para as aplicações de IA ao Direito.

A CEEUIA SJ volta-se a todos os desenvolvedores públicos e privados, responsáveis pela criação ou pela regulação e auditoria de instrumentos de IA para a eficiência da justiça. Estabelece primordialmente que todos os modelos criados devem ser orientados à concretização dos Direitos Humanos previstos na Convenção Europeia dos Direitos do Homem e na Convenção relativa à Proteção dos Dados Pessoais. Além disso, busca orientar as aplicações de IA produzidas em todas as áreas do direito, com especial cuidado ao âmbito penal, o qual, segundo a Carta, pode gerar efeitos discriminatórios mais acentuados.

A Carta estrutura-se da seguinte forma: cinco princípios fundamentais; Apêndice I, que traz um estudo aprofundado sobre a utilização da IA nos sistemas judiciais, nomeadamente as aplicações de IA que processam decisões e dados judiciais; Anexo II: Quais as utilizações da IA nos sistemas judiciais europeus?; Apêndice III: Glossário; Apêndice IV: Lista de verificação para integrar os princípios da Carta no método de processamento.

Os cinco princípios são os seguintes: respeito aos direitos fundamentais, não discriminação; qualidade e segurança; transparência, imparcialidade e equidade; controle do usuário. A Carta exige que os sistemas de IA criados para o Direito estejam em consonância com a proteção dos direitos fundamentais previstos na Convenção

Europeia dos Direitos do Homem (CEDH) e também nas regulações europeias de proteção de dados pessoais. Menciona ser essencial que, desde a concepção dos projetos de IA, seja incorporada a salvaguarda às garantias de “acesso ao juiz e do direito a um julgamento justo (igualdade de armas e respeito pelo processo contraditório) [...] respeito dos princípios do Estado de direito e da independência dos juízes no seu processo decisório” (UNIÃO EUROPEIA, 2018).

O princípio da não discriminação dá especial enfoque na prevenção da ocorrência de resultados que intensifiquem ou reflitam quaisquer discriminações contra grupos minoritários; além de requerer especial cuidado com tratamento de dados sensíveis. Eventuais resultados discriminatórios devem ser interrompidos e descontinuados ou corrigidos.

A grande preocupação com os dados usados também se reflete no princípio da qualidade e segurança, que pretende garantir projetos cujo manuseio seja de fontes qualificadas, rastreáveis e certificadas de dados processuais, os quais devem ser mantidos íntegros durante a execução. O tratamento dos dados deve ser realizado por equipes plurais (de culturas, raças e orientações diferentes e de áreas diversas), que dialoguem entre si e que sejam

[...] capazes de recorrer amplamente à experiência dos profissionais relevantes do sistema de justiça (juízes, procuradores, advogados, etc.) e pesquisadores/professores nos campos do direito e das ciências sociais (por exemplo, economistas, sociólogos e filósofos) (UNIÃO EUROPEIA, 2018).

O princípio da transparência, imparcialidade e equidade relaciona-se ao tema do segredo comercial. A Carta é lacônica ao estabelecer que para resolver o *tradeoff* entre explicabilidade e proteção ao segredo de negócio, a propriedade intelectual dos produtos de IA criados pelo âmbito privado deve ser balanceada com a transparência. Mas não identifica quais os critérios podem ser usados nesse balanceamento.

A inquietação aqui é com a governança e auditabilidade dos modelos pelos cidadãos e, para tanto, não se pode passar ao largo da inteligibilidade dos mesmos. A Carta coloca que a primeira opção técnica:

[...] é a transparência técnica total (por exemplo, código-fonte aberto e documentação), por vezes limitada pela proteção dos segredos comerciais. O sistema também pode ser explicado em linguagem clara e familiar (para descrever como os resultados são produzidos) comunicando, por exemplo, a natureza dos serviços oferecidos, as ferramentas que foram desenvolvidas, o desempenho e os riscos de erro. As autoridades ou peritos independentes podem ser encarregados de certificar e auditar métodos de processamento ou

de prestar aconselhamento prévio. As autoridades públicas poderiam conceder a certificação, que seria regularmente revista.

Sugere-se inclusive a criação, pelas autoridades públicas, de certificações técnicas de explicabilidade, o que ajuda no controle.

Por fim, o princípio do controle do usuário, que está intimamente relacionado ao princípio anterior, visa garantir que usuários (cidadãos e funcionários do sistema judicial) sejam informados em linguagem clara, acessível e inteligível sobre o uso de sistemas de IA. Além disso, aponta a necessidade da previsão de revisão e recorribilidade das decisões algorítmicas.

Todos os princípios dialogam com o conceito de ética **no, pelo e para** o *design* (Dignum, 2018), os quais exigem que desde a concepção e durante todo desenvolvimento seja realizado o desenho de limites ao uso da IA.

A CEEUIA SJ traz também importantes referenciais no Anexo I sobre o estado de utilização dos algoritmos de inteligência artificial nos sistemas judiciais dos Estados-Membros do Conselho da Europa; a panorâmica das políticas de dados abertos relativos às decisões judiciais nos sistemas judiciais dos Estados-Membros do Conselho da Europa; uso de IA em sistemas preditivos e em matéria penal. Um quadro interessante aponta uma lista de serviços já usados no sistema jurídico ao redor do mundo:

Quadro 4- Lista não exaustiva dos serviços jurídicos que recorrem à inteligência artificial na área jurídica.

Software	Estado	Tipo de
Doutrine.fr	França	Motor de busca
Prédice	França	Análise (exceto processos penais)
Análise da Jurisprudência	França	Análise (exceto processos penais)
JurisData Analytics (LexisNexis)	França	Motor de busca, Análise (exceto casos criminais)
Luminância	Reino Unido	Análises
Watson/Ross (IBM)	ESTADOS UNIDOS DA AMÉRICA	Análises
HART	Reino Unido	Análise (criminal, risco de reincidência)
Lex Machina (LexisNexis)	ESTADOS UNIDOS DA AMÉRICA	Análises

Fonte: União Europeia (2021).

O Anexo II classifica algumas utilizações em: “possíveis”, exigindo precauções metodológicas consideráveis (em disputas e métodos alternativos civis, litígios em linha, em infrações penais); “a considerar na sequência de estudos científicos suplementares” (perfilização de juízes, antecipação de decisões judiciais); “a serem considerados com as reservas mais extremas” (perfilização em matéria penal e normas baseadas em quantidade). Isso é, classifica a IA em níveis de risco por área de aplicação no Direito ou por tipo de resultado e inclui como resultado de risco médio a jurimetria e risco extremo a perfilização em matéria penal.

Há também um glossário para compreensão dos termos técnicos comumente utilizados nos projetos que envolvem IA e uma lista para auxiliar a averiguação de cumprimento da Carta em cada etapa de desenvolvimento desses projetos.

Inspirado na Carta Europeia de 2018, mas posterior a alguns projetos que já vinham sendo implantados no judiciário brasileiro (como o Projeto Victor), em 2020 o Conselho Nacional de Justiça (CNJ), publicou a Resolução n. 332, que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências.

Em modelo um pouco diverso da Carta, a Resolução traz 31 artigos e 14 considerandos. Os considerandos iniciais contextualizam a Resolução, ressaltando a relevância da IA para a Justiça brasileira, salientando a preocupação de que essa IA seja aplicada em respeito à proteção de dados pessoais e em consonância com a LGPD, o direito à privacidade e a proteção dos direitos fundamentais.

No Capítulo I encontram-se as disposições gerais, em modelo mais simples que a Carta Europeia, que dedica um Anexo para expor conceitos e produção científica sobre IA. Define no art. 3º, I, por exemplo, algoritmo como a “sequência finita de instruções executadas por um programa de computador, com o objetivo de processar informações para um fim específico” (BRASIL, 2020).

O Capítulo II, assim como a CEEUIA SJ, traz o princípio do respeito aos direitos fundamentais e ressalta a proteção aos dados pessoais, em especial, os dados sensíveis. Apesar de enunciar essa proteção especial e remeter à LGPD, não especifica de modo prático como cumprir o princípio, em especial quando estão envolvidos dados pessoais obtidos nos processos.

O princípio da não discriminação está previsto no Capítulo III e destaca que os sistemas de IA devem servir para minimizar preconceitos e discriminações,

promovendo julgamento justo. Nesse sentido, a Resolução determina que antes de os modelos de IA serem colocados em produção, devem ser homologados previamente para que se identifiquem vieses discriminatórios e, uma vez encontrados esses resultados é necessária adoção de medidas para correção. Entretanto, se não for possível a correção do modelo, esse deve ser descontinuado, registrando-se os motivos que motivaram a cessá-lo. A Resolução não aponta, todavia, como realizar essa homologação, nem tampouco sobre como serão homologados os modelos produzidos antes de 2020.

A publicidade e a transparência aparecem no Capítulo IV. A Resolução define e precisa no que consiste a transparência no art.8º,

Art. 8º Para os efeitos da presente Resolução, transparência consiste em:
 I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais;
 II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial;
 III – documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento;
 IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial;
 V – apresentação dos mecanismos de auditoria e certificação de boas práticas;
 VI – fornecimento de **explicação satisfatória** e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial (BRASIL, 2020, grifo nosso).

Nota-se que a Resolução buscou deixar claros pontos que devem ser atendidos pelos projetos de IA desenvolvidos para o Judiciário, para que tenham um resultado transparente. Apesar de não se referir ao termo exato “IA Explicável”, a Resolução traz no inciso IV a necessidade de explicação inteligível dos modelos aos cidadãos e jurisdicionados afetados. A questão de como enfrentar a opacidade dos modelos de IA é sempre um desafio tecnológico, principalmente nos modelos de aprendizagem profunda.

No caso da IA aplicada ao Direito, a explicabilidade, a transparência e o dever de “divulgar os métodos de processamento de dados, além de contribuir para o respeito pelo princípio do devido processo, excluiria – ou confirmaria – a existência de possíveis parâmetros discriminatórios” (FRONZA, 2021, p. 539). Mas o que seria o fornecimento de explicação satisfatória, prevista no art. 8º, VI?

Aqui cabe também voltar ao dispositivo 20 da LGPD, que traz a necessidade de revisão das decisões automatizadas “que afetem seus interesses, incluídas as decisões

destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade”. Estariam as decisões que utilizam algum tipo de tecnologia de IA para o Judiciário afetadas por este artigo, já que atingem aspectos da personalidade? Se assim for, cabe a pergunta: todas as decisões tomadas atualmente no Poder Judiciário que se utilizam de algum modelo de IA seriam recorríveis? Fato é: há que se pensar em novas previsões processuais recursais para os casos de decisões judiciais que se baseiam de alguma forma em IA.

O Capítulo V apresenta os princípios da governança e da qualidade. A governança algorítmica passa pela compreensão de que a produção de IA não é neutra e por isso necessita de regras éticas para manejo. Em especial quando aplicada ao setor público, a IA deve ter uso passível de controle democrático, eis que automatiza processos decisórios tomados por servidores públicos e gestores que afetarão os cidadãos (Silveira, 2017).

Com esse fim, a Resolução cria a Sinapses, plataforma para depósito de todos os modelos de IA desenvolvidos para o Judiciário brasileiro e que cria um padrão cooperativo entre as instâncias judiciais brasileiras, poupando esforços e economizando tempo, ao permitir a interação transparente e (re)uso dos resultados positivos. Além disso, prevê que os projetos devem ser devidamente informados desde o início ao CNJ, que destinará área própria de seu sítio eletrônico para divulgação dos modelos criados, os quais também deverão respeitar a interface de programação de aplicativos (API) estabelecida pelo CNJ.

O Capítulo VI prevê o princípio da segurança com preocupação e enfoque nos dados. Segurança da fonte (que deve ser preferencialmente governamental); proteção contra alterações, destruições, modificações extraviadas ou transmissões e acessos não autorizados do *dataset* e armazenamento seguro.

O princípio do controle do usuário está no Capítulo VII. Aqui se propõe a inteligibilidade dos sistemas de IA para o Judiciário, pretendendo garantir efetiva informação. A criação dos modelos deve proporcionar informação compreensível e que possibilite sua supervisão constante por cidadãos e magistrados.

A Resolução dedica o Capítulo VIII ao tratamento da pesquisa, do desenvolvimento e da implantação de serviços de Inteligência Artificial. Em consonância com o *design* ético, prevê que as equipes de desenvolvimento de IA para a justiça sejam plurais, diversas e interdisciplinares em todas as etapas do processo. O respeito aos direitos e garantias fundamentais também deve perpassar todo o projeto.

Em especial as pesquisas no âmbito da IA aplicada ao reconhecimento facial devem ser previamente autorizadas pelo CNJ.

No que tange aos modelos cuja aplicação destina-se ao âmbito do direito penal, a Resolução diz que sua produção não deve ser estimulada (principalmente se forem modelos preditivos). Mas o que seria exatamente “não estimular”? Permitir sem fazer alardes? Na sensível área penal há ainda uma polêmica exceção prevista nos §§ 1º e 2º:

§ 1º Não se aplica o disposto no caput quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.

§ 2º Os modelos de Inteligência Artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização. (BRASIL, 2020).

Isso é, a Resolução permite que sejam criados modelos para cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo sem, no entanto, tratar da polêmica relacionada à proteção da propriedade intelectual das empresas criadoras de modelos de IA, apontada por autores como Pasquale (2015) e O’Neil (2016). Nesse aspecto, traz apenas que, preferencialmente, os modelos devem ser criados com código aberto. Preferencialmente não é impedimento aos códigos fechados, mas permissão justificada.

Aqui também se deve relembrar que, como apontado no Capítulo 2 dessa pesquisa, a área penal demanda limitações mais rigorosas, já que é área sensível, que aparece de forma reiterada como preocupação ética da produção científica recente. Além de ser tema que mais aparece na produção científica recente em ética para IA aplicada ao Direito, no aspecto dogmático o Direito Penal sempre foi considerado “última *ratio*”, isso é, medida de punição mais extrema que lida com a restrição à liberdade de ir e vir e, por isso, que deve ser usado em último caso. O uso de IA nessa área ainda carece, portanto, de limitações mais precisas por parte das diretrizes regulatórias, tal qual é a Resolução n. 332/CNJ.

O Capítulo IX traz o princípio da prestação de contas e da responsabilização que busca trazer parâmetros para transparência de valores e custos dos projetos, nomes dos responsáveis e envolvidos, resultados planejados. Destaca-se aqui a previsão do art. 26:

Art. 26. O desenvolvimento ou a utilização de sistema inteligente em desconformidade aos princípios e regras estabelecidos nesta Resolução será objeto de apuração e, sendo o caso, punição dos responsáveis.

Tal artigo prevê a punição dos responsáveis pelos projetos que desobedeçam aos ditames apresentados na Resolução, após apuração, sem especificar, entretanto, como se dará o procedimento que poderá gerar essa responsabilização, qual tipo de responsabilidade e como aplicá-la.

O último Capítulo (X) apresenta as disposições finais que informam a aplicação imediata da Resolução, inclusive aos projetos em andamento.

A partir da análise pormenorizada de cada um dos instrumentos normativos, criou-se a tabela abaixo, que busca apontar quais são os indicadores que surgem em comum nas diretrizes comparadas e quais os desafios essas normativas enfrentam, no que tange a IA aplicada ao Direito:

Quadro 5 - Indicadores e desafios comuns às diretrizes que regulam IA aplicada ao Direito

INDICADOR NORMATIVO	DESAFIO	RESOLUÇÃO 332/CNJ	CEEUIA SJ
Respeito aos direitos fundamentais	Imprecisão do termo; interpretabilidade	Capítulo II	Art. 1º
Não discriminação	Como avaliar resultados discriminatórios?	Capítulo III	Art. 2º
Qualidade e segurança	Contratação de equipes plurais para o desenvolvimento dos projetos e seleção criteriosa da base de dados.	Capítulos V e VI	Art. 3º
Transparência	Criação de certificações que norteiem a avaliação de inteligibilidade dos modelos.	Capítulo IV	Art. 4º
Controle do usuário	Criação de normas processuais de recorribilidade para os casos de decisões judiciais (e extrajudiciais) que utilizam IA.	Capítulo VII	Art. 5º

Classificação em riscos	Como lidar com cada um dos riscos classificados? É necessário proibir certos usos no Direito? O que significa não estimular?	Art. 23	Anexo II
Glossário	Precisar os termos.	Capítulo I	Apêndice III

Nota-se que as iniciativas trazem balizas importantes, que servem de guia geral para todo o desenvolvimento e implantação de modelos de IA aplicada ao Direito – desde a concepção até o acompanhamento da evolução. Ambas trazem como indicadores normativos comuns os princípios do respeito aos direitos fundamentais, da não discriminação, da qualidade e segurança, da transparência e do controle do usuário. Embora se reconheça o alto risco em se utilizar IA na área do Direito Penal, há uma certa dificuldade em estabelecer os limites à sua utilização. Trazem também em comum glossário de termos que sinaliza a preocupação em precisar e explicitar o vocabulário usado na tecnologia para o público específico que liga com esses projetos no Direito.

São, portanto, diretrizes genéricas, no sentido de que apresentam mais princípios gerais e menos regras práticas. Ambas não trazem sanções, mas a brasileira prevê a possibilidade de punição (sem identificar, no entanto, como e por quem essa punição será efetivada).

Para além disso, foi possível observar que as diretrizes principiológicas geram desafios comuns quando se aplica IA ao Direito, como compreender o que é uma aplicação de IA que respeite os direitos fundamentais; quais os parâmetros para interpretabilidade aceitável; como avaliar resultados discriminatórios; a necessidade de contratar equipes plurais para o desenvolvimento dos projetos e seleção criteriosa da base de dados e de criação de certificações que norteiem a avaliação de inteligibilidade dos modelos. Com o crescimento do uso de aprendizado de máquina no Direito, impõe-se a criação de normas processuais de recorribilidade para os casos de decisões judiciais (e extrajudiciais) que utilizam IA e que direcionem como lidar com cada um dos riscos classificados da IA no Direito.

4 INPUTS ÉTICOS APONTADOS PELOS GRUPOS DE ATORES ENVOLVIDOS COM INTELIGÊNCIA ARTIFICIAL

A palavra ética pode suscitar algum desconforto, principalmente quando usada no âmbito da tecnologia. Isso porque pode aludir à ideia de limitação ou trava ao desenvolvimento tecnológico. Reitera-se que essa pesquisa utiliza o termo ética por entender, como Boddington (2017), que as questões éticas de agir e de ser de cada sociedade têm relevante impacto na imposição e produção normativa.

Nesse sentido, “a natureza normativa da ética significa que simplesmente descrever a maneira como as pessoas agem não dará conta da ação ética” (BODDINGTON, 2017, p. 9, tradução nossa). Isso é, a ética pode ser entendida – e é essa a compreensão dessa pesquisa – como algo que indica, determina e esculpe a normatividade e as imposições de limites às ações dos indivíduos sociais, bem como os valores fundamentais a serem considerados em cada cultura.

Parte-se, assim, do pressuposto de que o grande desafio está em avaliar os impactos produzidos pela IA na sociedade e, nesse sentido, buscar quais as estratégias mais eficazes para garantia dos resultados éticos dela advindos, bem como quais os efeitos ou externalidades negativas podem ser considerados como produtos eticamente inadmissíveis.

Dessa forma, a pesquisa pretende a seguir averiguar a percepção dos limites éticos à IA, dos atores que de algum modo se relacionam ao tema. Busca-se, com isso, compreender como esses atores enxergam seu papel no desenvolvimento e criação de critérios de limitação desses projetos de IA e qual sua visão sobre como deve ser uma atuação ética.

4.1 METODOLOGIA DE INVESTIGAÇÃO PARA AVERIGUAÇÃO DA PERCEPÇÃO DOS ATORES SOCIAIS

Para averiguar a percepção sobre limites éticos à IA dos sujeitos que de algum modo se relacionam ao tema, foram utilizadas entrevistas semiestruturadas. Isso é, entrevistas com perguntas pré-definidas, mas com liberdade para a entrevistadora conduzir a entrevista de forma a adicionar questões e cuja intenção é colher as

compreensões dos indivíduos acerca das suas experiências no assunto (Colognese & Melo, 1998).

A realização das entrevistas foi instrumento importante para cumprir a finalidade proposta neste capítulo, pois serviu para “mapear práticas, crenças, valores e sistemas classificatórios de universos sociais específicos, mais ou menos bem delimitados, em que os conflitos e contradições não estejam claramente explicitados” (DUARTE, 2004, p. 215). Permitem, assim, estimular as manifestações dos atores acerca de suas experiências, a fim de auxiliar na interpretação do fenômeno objeto de estudo.

4.2 SELEÇÃO DOS ENTREVISTADOS

Tendo em vista o objetivo da pesquisa foi necessário selecionar entrevistados dentre grupos sociais representantes dos diversos segmentos envolvidos de alguma maneira com IA e que, portanto, poderiam interagir de modo a contribuir com a discussão da pesquisa e com o fenômeno estudado. Foram, então, definidos três grupos de entrevistados: a) profissionais das áreas da tecnologia de informação, ciência da computação e engenharia de *software* que trabalham ou trabalharam em projetos de aplicação de IA (Grupo 1); b) magistrados e ministros, que lidam com o desenvolvimento de IA no Poder Judiciário brasileiro (Grupo 2) e c) cientistas sociais que pesquisam as consequências éticas da aplicação de IA (Grupo 3).

De cada segmento foram selecionados três a cinco entrevistados. Embora a pesquisadora tenha se esforçado para cumprir a paridade de gênero e raça, cabe destacar a dificuldade em encontrar, em especial no grupo da tecnologia, pesquisadoras mulheres e negras.

Dentre os profissionais da área de tecnologia de informação, ciência da computação e engenharia de *software* que trabalham ou trabalharam em projetos de aplicação de IA (Grupo 1), foram selecionados três entrevistados: um integrante do Laboratório de Inteligência Artificial da Universidade de Brasília (AILAb); um integrante do Laboratório de Inovação do Tribunal Federal da 3ª Região (ILabTRF3) e uma cientista da computação, pesquisadora, *hacker* antirracista e membro do conselho consultivo de segurança do *TikTok* no Brasil. Os contatos foram realizados por meio das redes sociais e *e-mail*.

No grupo de atores do Poder Judiciário brasileiro (Grupo 2), foram selecionados quatro magistrados: uma magistrada de primeiro grau; um desembargador de tribunal de

justiça estadual; o Ministro Presidente do Superior Tribunal de Justiça e o Ministro Presidente do Supremo Tribunal Federal. Optou-se, para formar um grupo de entrevistados plural, por selecionar um representante de cada grau de Jurisdição. Os contatos foram realizados por meio dos canais oficiais disponibilizados no *site* dos respectivos tribunais aos quais os juízes são vinculados.

Por fim, selecionaram-se quatro atores do grupo dos cientistas sociais que pesquisam as consequências éticas da aplicação de IA (Grupo 3). Uma professora de Direito e Tecnologia da UERJ; uma professora de Direito do IDP e membro da ANPD; um professor de sociologia e especialista em teoria da inovação e sociologia econômica da USP e um professor da sociologia da Universidade Federal do ABC e coordenador do grupo Tecnopolítica.

A escolha foi feita considerando a existência de vieses de cada grupo representado e dos possíveis interesses envolvidos.

Os contatos foram realizados com cada entrevistado por meio das redes sociais e *e-mail*.

No total foram realizados, portanto, um universo de onze entrevistas (três profissionais da área de tecnologia e quatro magistrados que atuam em projetos de IA ao Poder Judiciário e quatro pesquisadores das ciências sociais).

Quadro 6– Perfil dos entrevistados

	Descrição do grupo	Atuação dos participantes	Total de entrevistados
GRUPO 1	Profissionais das áreas da tecnologia de informação, ciência da computação e engenharia de software que trabalham ou trabalharam em projetos de aplicação de IA	<ul style="list-style-type: none"> • E1: Um integrante do Laboratório de Inteligência Artificial da Universidade de Brasília (AILab); • E2: Um integrante do Laboratório de Inovação do Tribunal Federal da 3ª Região (ILabTRF3) • E3: Uma cientista da computação, 	3

		<p>pesquisadora, hacker antirracista e membro do conselho consultivo de segurança do TikTok no Brasil</p>	
GRUPO 2	Atores do Poder Judiciário brasileiro	<ul style="list-style-type: none"> • E4: Uma magistrada de primeiro grau; • E5: Um desembargador de tribunal de justiça estadual; • E6: O Ministro Presidente do Superior Tribunal de Justiça • E7: O Ministro Presidente do Supremo Tribunal Federal. 	4
GRUPO 3	Cientistas sociais que pesquisam as consequências éticas da aplicação de IA	<ul style="list-style-type: none"> • E8: Uma professora de Direito e Tecnologia da UERJ; • E9: Uma professora de Direito do IDP e membro da ANPD; • E10: Um professor de sociologia e especialista em teoria da inovação e sociologia econômica da USP • E11: Um professor da sociologia da Federal do ABC e 	4

		coordenador do grupo Tecnopolítica.	
--	--	-------------------------------------	--

Fonte: elaborado pela autora

4.3 ASPECTOS ÉTICOS DA PESQUISA

Em atendimento ao que estabelece a Resoluções nº. 466 de 12 de dezembro de 2012 que determina: “As pesquisas envolvendo seres humanos devem atender aos fundamentos éticos e científicos pertinentes”; e ainda a Resolução nº 510 de 7 de abril de 2016 (BRASIL, 2016) que dispõe sobre as normas aplicáveis às pesquisas que envolvam dados obtidos por participantes ou informações identificáveis ou que possam acarretar riscos maiores do que aqueles da vida cotidiana, esta pesquisa foi submetida ao Comitê de Ética em Pesquisa da UNB via Plataforma Brasil e devidamente aprovada. O número do Certificado de Apresentação de Apreciação Ética (CAAE) é 31920120.1.0000.5540.

4.4 ROTEIRO E APLICAÇÃO DAS ENTREVISTAS

Foram realizadas entrevistas individuais, com duração de em média uma hora. O roteiro de entrevistas semiestruturadas foi elaborado levando em conta os objetivos e os temas relevantes da pesquisa, norteadores das questões.

Assim, foram realizadas as seguintes questões aos entrevistados(as):

1) Gostaria que comentasse sobre a sua experiência na área de regulação de dados e IA. Ou no uso da IA no Direito.

2) Com base na sua experiência, como o(a) senhor(a) avalia o uso dessa ferramenta atualmente no Brasil?

3) Como o(a) Senhor(a) avalia a questão ética relacionada à aplicação de IA no Direito? Poderia relatar?

4) Na sua atuação o Senhor (a) já se deparou com alguma questão ética na aplicação de IA? Como foi resolvida?

5) Como o(a) senhor(a) avalia diretrizes existentes para IA?

6) Considerando sua experiência e desafios éticos apontados, o que o(a) Senhor(a) compreende sobre

- a) O uso de decisões automatizadas no Direito
- b) O uso de dados pessoais (sensíveis ou não) na produção de IA?
- c) Transparência e explicabilidade
- d) Os vieses
- e) A possibilidade de *Open Justice*?

7) O(a) Senhor(a) teria alguma sugestão/vê algum caminho para melhorar o controle ético da IA aplicada ao Direito? Ou deseja acrescentar algo?

Algumas perguntas foram alteradas de acordo com o papel do ator entrevistado e quando necessário, foram acrescentadas por: “pode falar um pouco mais sobre isso?”, “Por que isso?”, “Explique isso”, “Deseja acrescentar algo mais ao encerrar a entrevista?”.

4.5 ANÁLISE DOS DADOS: “O MEU PAÍS É O MEU LUGAR DE FALA”

Os dados obtidos nas entrevistas foram examinados por meio de análise de conteúdo, ou seja, “uma técnica de investigação que através de uma descrição objectiva, sistemática e quantitativa do conteúdo manifesto das comunicações, tem por finalidade a interpretação dessas mesmas comunicações” (BARDIN, 1977, p. 36)⁷¹.

A análise compõe-se de três fases, a saber: a) pré-análise, por meio da qual se organizam os dados coletados para análise e se criam as hipóteses e indicadores dos documentos; b) exploração do material, em que se criam categorias e identificação de unidades de significação; c) tratamento dos resultados, inferência e interpretação, em que se dão as interpretações (VERGARA, 2011).

Em resumo, o que se buscou com a análise de conteúdo nessa pesquisa foi codificar os conteúdos significativos para formar um conjunto comum nas narrativas apresentadas nas entrevistas o qual expressasse questões éticas que permeiam a IA aplicada ao Direito. Valendo-se, portanto, do discurso das entrevistas e após inúmeras e minuciosas releituras, foram selecionadas unidades de análise e categorizadas as respostas. Isso é, foi possível identificar padrões nas entrevistas, valores, princípios éticos, interesses de cada grupo selecionado.

⁷¹ Sampaio e Lycarião (2021) realizaram um mapeamento bibliográfico de 3.484 artigos no Brasil que afirmavam utilizar como metodologia a análise de conteúdo. Nas conclusões, defendem que as pesquisas devem ir além do tradicional manual de Bardin (1977), o qual tem sido utilizado de forma frequente no Brasil (quase como um ciclo vicioso) e como autora exclusiva da técnica de análise. Essa tese utiliza a autora reconhecendo as limitações de seu manual, assim como a crítica de Sampaio e Lycarião.

A partir da amostra das respostas, a categorização foi criada de duas formas: a) por correlação das respostas dos entrevistados ao marco teórico, o que possibilitou a interpretação dos dados de forma a obter inferências válidas e replicáveis (SAMPAIO; LYCARIÃO, 2021; KRIPPENDORFF, 2018) e b) correlacionando as respostas a algo que aludisse a algum conteúdo cultural brasileiro. A escolha de categorizar com base em elementos culturais brasileiros justifica-se em primeiro lugar, para conectar as respostas com questões multidisciplinares (que fugissem do campo estrito da academia, mas que se conectassem de alguma forma com a temática presente nas respostas) e em segundo lugar, para trazer subsídios que se conectassem a realidade brasileira e fossem próprios do Brasil, país onde se realizou a pesquisa, ao invés de importar elementos estrangeiros (realizando narrativa inspirada em Valter Hugo Mãe, que traz em sua obra “As doenças do Brasil” questões profundas das relações entre colonizador e colonizado). Afinal, como cantou Elza Soares, “o meu país é o meu lugar de fala”.

Com base nas etiquetas criadas a partir do conteúdo qualitativo das entrevistas, buscou-se compreender como os conteúdos significativos para a pesquisa foram tratados nas respostas dos atores. Foram elaboradas, então, seis categorias e quando se viu necessidade (quando se encontrou padrões diversos nas entrevistas que deviam ser reportados), também foram concebidas subcategorias, como se pode observar no Quadro 7.

Quadro 7– Categorias, subcategorias e autores de referência.

CATEGORIA	SUBCATEGORIA	AUTOR(ES)
IA no Direito	Momento Inescapável	Walker (2021); Susskind (2017; 2019); CUEVA (2021); Kahneman; Sibony; Sunstein (2021)
	IA como “Coringa Intelectual”	Benjamin (2020); Jordan (2018)
Pretensões principiológicas	-	Burle; Cortiz (2019); Floridi e Cowls (2019); Kaufman (2021); Mulholland; Frajhof (2021)
“Desses que vivem no escuro em plena luz”	Transparência e explicabilidade: sentido multidisciplinar	Miller (2019); Wachter, Mittelstadt; Floridi (2018)

	Opacidade e <i>Tradeoffs</i>	Barredo Arrieta et. al. (2020); Pasquale (2015); Frazão e Goettenauer (2021); Frazão (2021)
Dadaísmo judicial	-	Wimmer (2021); Mulholland; Frajhof (2019); Junquillo; Maia Filho (2021); Becker; Torres (2021)
Vieses como “limonada pronta”	-	Kahneman; Sibony; Sunstein (2021); Sunstein (2019); Birhane (2021); Silva (2021); O’Neil (2016)
“E agora José?”	-	Floridi et al. (2018); Dignum (2018); Boddington (2017).

Fonte: elaborado pela autora

Cabe destacar, antes de iniciar a análise, que se optou por eliminar da transcrição das entrevistas eventuais erros e vícios de linguagem, para que a leitura fique mais fluída.

4.5.1 IA no Direito

Na primeira categoria buscou-se conectar as questões colocadas pelos participantes a respeito do uso e aplicação da IA no Direito, às perspectivas do marco teórico. A partir da análise das respostas, surgiram duas subcategorias: “momento inescapável” e “há necessidade?”, as quais serão debatidas a seguir.

4.5.1.1 Momento Inescapável

A maioria dos entrevistados avaliou o uso da IA no Direito como um “momento inescapável”, na esteira do que defendem Walker (2021), Susskind (2017; 2019) e Cueva (2021).

Dois dos três entrevistados do Grupo 1, compreenderam que a aplicação de IA ao Direito, principalmente no Brasil, mais que uma realidade é uma necessidade premente em especial nas tarefas repetitivas e em razão da enorme quantidade de dados processuais existentes. O E1 descreveu o projeto que desenvolve atualmente como um

[...] Sistema de apoio à construção de minutas de votos para os gabinetes [...] pegar processos e agrupar processos. Isso é uma tarefa reiteradamente solicitada pelos tribunais e pelos órgãos... Eu acho que ela (IA) é uma **necessidade premente**. Não tem como a gente caminhar no sistema judiciário brasileiro com essa quantidade exorbitante de processos que nós temos, não só na base, mas também nas cortes superiores. Sem esse encaminhamento de automação, de reconhecimento de padrão, de juntada de processos similares, levantamento de jurisprudência com relação com precedentes isso, sem dúvida vai trazer uma aceleração. Uma maior precisão nos trabalhos de análise de processos. (E1, grifo nosso)

O E2 descreveu que os modelos que atualmente desenvolve servem para

[...] Garantir mais celeridade, evitar decisões conflitantes, garantir segurança jurídica, mas sem tirar a autonomia do servidor e do magistrado, que são eles que selecionam e eles que fazem a avaliação disso daí [...] nosso foco hoje com os dois principais projetos nossos [...] são [...] melhorar a celeridade e segurança jurídica. (E2)

As percepções corroboram o que Susskind (2019; 2017) define como segunda onda de Inteligência Artificial para o Direito: demanda do judiciário e do mercado pela mudança no formato de entrega dos serviços jurídicos aos cidadãos, a partir de manipulação de grande massa de dados e aprendizado de máquina. A utilização de algoritmos torna-se, segundo o autor, imperiosa para identificação de padrões e predições. Sua utilização, como apontam as respostas, são consideradas como “necessidade premente” para um sistema judicial moroso e sobrecarregado como o brasileiro.

No Grupo 2, a E4 observou:

[...] é uma ilusão a gente achar que vai conseguir dar conta de todas essas etapas dos processos judiciais sem o auxílio, sem o apoio da tecnologia. O quanto da rotina do judiciário é consumida hoje por uma atividade mecânica de baixa atividade cognitiva. A inteligência artificial entrou e muito da automação entrou no judiciário por esses motivos: substituição de tarefas mecânicas feitas por uma automação que não necessariamente utiliza inteligência artificial e por uma questão de necessidade: é um volume muito grande de processos. (E4)

O E5 apontou

[...] eu acho que a ferramenta é essencial, extraordinária, principalmente para mostrar alguns padrões de decisão que às vezes escapam ao julgador e pode também racionalizar o trabalho em demandas repetitivas. Pode ser bem possível atribuir ao demandante a responsabilidade de demonstrar por que o caso dele é diferente daquele caso julgado da jurisprudência. (E5)

Ressaltou ainda que o uso de IA no Judiciário pode representar uma diminuição na carga de trabalho dos magistrados:

[...] significa que o juiz vai ter mais tempo para os casos difíceis e complexos principalmente envolvendo demandas estruturais e litígios complexos ou também outros casos que são individualizados e que você não consegue uma reprodução na jurisprudência por ter uma riqueza fática muito grande. (E5)

O E6 avalia que o uso da IA possibilitou

[...] melhora na triagem. Nos permitiu, na verdade, prestar um serviço jurisdicional mais rápido, mais coerente e um custo menor para a sociedade, pois vem a possibilitar a distribuição de servidores para o desempenho de atividades intelectuais, atualmente inviáveis de serem desempenhadas por algoritmos inteligentes. (E6)

O E7 expôs as

[...] vantagens da IA em três planos: (I) para a visão macrossistêmica do dia-a-dia dos Tribunais, considerando variadas circunstâncias geográficas, estruturais e orçamentárias; (II) para a jurisdição constitucional, ante o poder-dever de uniformização dos entendimentos; e (III) para o fortalecimento da Vocaç o Constitucional do STF, o que implica racionalizaç o das compet ncias recursais. (E7)

E complementou entender que

[...] a an lise preditiva e o agrupamento de demandas repetidas servem para o tratamento ison mico de situaç es subjetivas cuja natureza j  foi pacificada pelo Supremo Tribunal Federal. A IA   nossa aliada na consecuç o de um dos objetivos cardeais do C digo de Processo Civil de 2015 – a jurisprud ncia s lida e coesa, m xime a observ ncia dos precedentes qualificados. (E7)

O que se nota em comum nas respostas do G2   o fato de salientarem os benef cios do uso da IA tanto em demandas repetitivas para desafogar o Judici rio, como tamb m para consolidar precedentes judiciais. A utilizaç o de aprendizado de m quina no Direito serviria, por essa perspectiva, em um primeiro momento, para apontar decis es de casos id nticos destoantes e, ap s, como redutor de ru dos, propiciando a higiene (correç o e limpeza) e organizaç o das decis es de casos corriqueiros em relaç o aos casos complexos (KAHNEMAN; SIBONY; SUNSTEIN, 2021).

Cabe ressaltar que a subcategoria “momento inescap vel” foi extra da da perspectiva apontada pela E8 do Grupo 3:

[...] Então eu acho que a gente tá em um **momento** que é um pouco **inescapável** em termos de desenvolvimento tecnológico e de utilização de sistemas de inteligência artificial; que a gente chegou no momento histórico em que de fato nós temos resultados bastante práticos e bastante evidentes de que a utilização de sistemas de IA traz algum resultado, que é evidentemente mais rápido e mais eficiente, pelo menos em algumas áreas mais eficiente do que é a análise humana [...] por exemplo, nos sistemas que fazem a avaliação prévia de admissibilidade de recursos, como é o caso do STF, caso de alguns tribunais regionais, eu vejo que a aplicação desse sistema vai trazer um benefício gigantesco para a sociedade, porque, de fato, vai agilizar o processo judiciário, sem dúvida alguma. (E8)

No mesmo sentido o E10 pontuou que “existem decisões muito rotineiras, muito básicas, elementares, que não tem porque você precisar de advogado”. (E10)

A posição de que o funcionamento da Justiça brasileira é propício, por natureza e formato, à aplicação de tecnologias tais quais os modelos de AM, que exigem ampla base de dados, é salientada no prefácio à obra de Joshua Walker (2021, p. 10), que chega a afirmar, inclusive “que o Brasil poderia se tornar o principal centro de IA jurídica no mundo”. A “hiperjudicialização”, característica do sistema brasileiro, é circunstância que propicia e favorece a concretização de projetos que dependem de dados para reconhecimento e reprodução de padrões (CUEVA, 2021).

4.5.1.2 IA como “Coringa Intelectual”

O pesquisador de Berkeley, Michael Jordan (2018), alerta que “a inteligência artificial (IA) é o mantra da era atual”, mas deve ser encarada como uma disciplina da engenharia em ascensão que apresenta graves desafios à humanidade. Mais ainda: que os sistemas de IA têm sérias limitações e consequências e que o termo tem sido utilizado como uma espécie de “Coringa Intelectual” em instituições que desenvolvem o aprendizado de máquina.

Corroborando com esse entendimento, a E3 do Grupo 1 trouxe ao debate resposta que levou à criação da subcategoria “IA como “Coringa Intelectual”, quando afirmou:

[...] Acho que está havendo um certo despreparo em normatizar essas tecnologias. As pessoas estão colocando a responsabilidade da solução dos problemas que informam à sociedade em uma única tecnologia com uma visão muito limitada das possibilidades que nós temos de resolução. Tem uma preocupação: o mercado está acelerando o uso dessas tecnologias sem ter antes uma reflexão se realmente é necessário. Então eu diria que eu penso muito assim: é necessário ter IA em tudo que nós estamos vivendo hoje?

Porque a gente acaba ficando muito dependente do que essas tecnologias vão nos retornar, do que esse uso pode nos retornar e isso está mudando muito rápido o impacto e as nossas relações e a gente não está conseguindo acompanhar essas mudanças. (E3)

A resposta é um alerta que também vai ao encontro da visão de Benjamin (2021), para quem a tecnologia pode servir de “profilaxia moral”, cujos objetivos benevolentes colocados para possibilitar a inovação, podem esconder novas formas de controle social, racial e étnico.

Há que se ter claros os propósitos da inserção dessa tecnologia no cotidiano judicial, em especial para que não se reforcem desigualdades e vulnerabilidades. E também para que não se crie a falsa ilusão de que essa tecnologia é capaz de, sozinha, resolver os problemas da cultura morosa do sistema jurídico brasileiro. As finalidades e justificativas dos projetos de IA, nesse sentido, devem ir além do raso argumento de promoção de inovação pela inovação.

4.5.2 Ética na IA como pretensões principiológicas

A segunda categoria, “ética na IA como pretensões principiológicas”, traduz as respostas dos entrevistados à questão das normativas e diretrizes éticas da IA aplicada ao Direito. Os respondentes de todos os grupos foram unânimes na compreensão de que o que existe atualmente é um conjunto diverso e amplo de princípios, mas que ainda não há concordância sobre normativas efetivas, sancionatórias e que atribuam responsabilidades.

Como afirmou a E9

[...] na verdade a gente não tem ainda um consenso global sobre essas diretrizes. A gente vê muitos esforços, esforços fragmentados, uma verdadeira disputa entre diferentes organizações internacionais, mas ainda estamos distantes, eu diria, de um consenso. Tem alguns termos que a gente vê sendo utilizados muito frequentemente, associados à transparência, à explicabilidade, à ideia de uma IA centrada no ser humano, mas me parece que ainda é preciso avançar muito para dar mais concretude para esses princípios que são formulados ainda de maneira muito abstrata, muito ampla e que nas diferentes áreas da aplicação de IA talvez haja desafios próprios a serem enfrentados. Então, por exemplo, a ética da IA na saúde possivelmente terá um vetor, um conjunto de elementos orientadores da aplicação de IA por exemplo em indústrias, IA associada ao mercado de trabalho. Então eu acho que falar de ética em IA hoje é um grande guarda-chuva que talvez ainda precise ser melhor preenchido, melhor explorado”. (E9)

Essa fala se repete no respondente E8

[...] hoje em 2021 há cerca de 11 projetos legislativos tanto na Câmara quanto no Senado [...] são projetos que têm dois pontos comuns. O primeiro que é: são projetos conceituais, então, trazem conceitos sobre inteligência artificial, *machine learning*, algoritmos, interferência humana e às vezes conceitos inclusive ou amplos demais ou equivocados. Isso é um problema. Outro ponto em comum é o fato de que esses projetos têm uma natureza principiológica então são projetos de lei que não têm a pretensão de uma regulação normativa estrito senso sancionatória, mas são projetos de lei que têm uma principiologia muito semelhante ao que existe, por exemplo, aplicado a valores éticos [...] e aí a minha crítica é: essas são leis que são interessantes porque estão atentas ao que está acontecendo, mas elas são muito pouco efetivas ou eficientes no sentido de que elas repetem na realidade aquilo que já é aplicado concretamente [...] que os desenvolvedores e aplicadores dos sistemas já vêm tentando atender de uma certa forma. (E8)

De fato, em estudo de 2019, Burle e Cortiz mapearam princípios da IA em diretrizes mundiais divididos por seis dimensões⁷² e encontraram divergências entre os documentos, imprecisão conceitual e descrições genéricas sem determinações procedimentais sobre aplicação prática. Floridi e Cowls (2019), também fizeram esforço semelhante de pesquisa que pretendeu unificar em cinco os princípios aplicáveis a IA.

Os respondentes E10 e E11 ainda ressaltaram a importância de se ir além da implantação de uma IA ética, por meio da criação de regras que regulem os setores de modo a impor sérias consequências ao seu descumprimento. Em consonância ao que conclui Kaufman (2021), a principiologia existente atualmente nos diversos documentos e diretrizes mundiais não é suficiente para limitar de forma precisa as externalidades negativas advindas da IA.

Destacam-se aqui os seguintes trechos do E10

[...] aqui no Brasil tem várias iniciativas [...] que fazem afirmações genéricas [...] em alguns momentos tentam reproduzir princípios já definidos pela União Europeia. (E10)

E sobre o PL n. 21/20, avaliou:

[...] é genérico e ao mesmo tempo ela não define consequências. Então você não tem punições e não tem como saber o que acontece caso as recomendações ou as diretrizes não sejam atendidas. Esse é um dos problemas. Essa questão da legislação é um dos problemas básicos para você montar um grande e robusto ecossistema de inteligência artificial no Brasil que nós não temos [...] Então aqui a questão do Marco legal Inteligência

⁷² As dimensões foram agrupadas em: 2 Equidade (Fairness); 3 Confiabilidade e Segurança (Reliability & Safety); 4 Impacto Social (Social Impact); 6 (Sic) Responsabilidade (Accountability); 7 Privacidade & Segurança (Privacy & Security); 8 Transparência (Transparency).

Artificial é uma questão chave é construir um ambiente saudável que procure não garantir, garantir é uma palavra muito forte, mas que procure minimizar os impactos negativos e que a gente consiga trabalhar a ideia de uma inteligência artificial para todos, uma inteligência artificial voltada para a prosperidade, uma inteligência artificial que seja ética, que tenha parâmetros saudáveis e que não passe por cima de direitos básicos das pessoas, nem mesmo do direito das instituições ou de governo. Isso é superimportante em um mundo como da inteligência artificial que tem características muito próprias. A gente se esquece, mas concretamente são poucos países que expandem a fronteira na inteligência artificial [...] E isso do ponto de vista da ética e da legislação é fundamental porque muitas vezes a legislação se volta para controlar e castrar aos mais frágeis, quando o problema não está nos mais fracos, o problema está exatamente nos mais fortes! Em empresas e corporações muito poderosas. (E10)

O respondente E11 afirmou

[...] existe, portanto, o primado ético. Várias discussões éticas que eu acompanho. Mas não existe uma regulação [...] grandes empresas de tecnologia da informação, grandes plataformas que utilizam a inteligência artificial e a desenvolvem têm sempre [...] áreas que desenvolvem a discussão ética. Então eu não conheço nenhuma empresa agora que é contra a aplicação ética na inteligência artificial, mas conheço muitas que são contra a regulação. (E11)

Além disso, os respondentes do G1 manifestaram ainda que a existência apenas de princípios gera certo temor, receio e insegurança aos desenvolvedores de IA. O E2, por exemplo, colocou

[...] A gente tenta fazer dessa forma mesmo que às vezes isso demore um pouco mais, mas que tenha um rigor formal para a gente ter certeza de que tudo que a gente está fazendo é legal, é ético e vai colaborar, vai melhorar o serviço que o poder judiciário está prestando e não piorar. Porque é uma questão muito delicada ele está mexendo com coisas muito delicadas em todos os sentidos aqui. (E2)

Em resumo, o que os entrevistados compreendem é que falta concretude. A despeito da existência de variados documentos internacionais sobre ética e princípios aplicáveis à IA, não há ainda regulação que assegure os rumos de um desenvolvimento seguro para todos. Nesse sentido, temem os técnicos, estudiosos e cidadãos, os possíveis resultados e consequências das aplicações dessa tecnologia. O tema e sua complexidade demandam “esforço criativo” e pesquisas como essa, que pretendam suprir essa premente necessidade de codificação (MULHOLLAND; FRAJHOF, 2021).

4.5.3 “Desses que vivem no escuro em plena luz”

Em seu álbum mais recente, “Meu Coco”, produzido na pandemia da COVID-19, Caetano Veloso canta na música “Anjos Tronchos” sobre a era dos algoritmos. Um dos trechos da música inspirou a criação da categoria “desses que vivem no escuro em plena luz”, que representa as respostas obtidas acerca da transparência e explicabilidade nos resultados obtidos com a utilização de sistemas de IA.

4.5.3.1 Transparência e explicabilidade: sentido multidisciplinar

Há uma preocupação apontada por todos os atores entrevistados quanto à inteligibilidade dos modelos algorítmicos, em especial usados no Direito. Isso fica claro na resposta do E4 do G2:

[...] os algoritmos são geralmente opacos. A gente sabe a regra de aprendizagem, se existe a forma de um algoritmo não programado a gente sabe a regra de aprendizagem, mas você acaba que perde um pouco controle sobre essas etapas que vão acontecendo até chegar no resultado porque existem as chamadas *Black Box Operations*, ou seja, aquelas operações de caixa preta que dão opacidade a todo o desenvolvimento do início até o fim dele chegar a sugerir uma decisão para o servidor e para o juiz. E às vezes uma IA não é explicável nem mesmo para o próprio programador. Como a gente já sabe em razão dessa opacidade é difícil a gente conferir 100% de transparência àquela decisão judicial que se utilizou de alguma forma de um sistema de inteligência artificial. (E4)

Foi unânime, portanto, o reconhecimento da urgência da promoção do debate sobre transparência e explicabilidade dos sistemas de IA. Entretanto, foi comum aos respondentes o incômodo em não saber como concretizar uma IA suficientemente transparente e explicável ao ser humano. Para a E9 do G3

[...] a ideia de transparência e de explicabilidade surge de certo modo como uma reação a essa percepção de perda do controle ou de perda do elemento de causalidade, que é uma característica que a gente tradicionalmente sempre valorizou nas pesquisas científicas e que agora de certo modo se perde, visto que essas grandes bases de dados viabilizam a identificação de correlações ou a realização de inferências sem que qualquer relação de causalidade seja também explicitada e daí que acho que quando a gente começa a ver sistemas de IA em campos do poder público, no *score* de crédito, na tomada de decisão no Poder Judiciário, isso suscita uma preocupação grande porque existem certos valores que são muitos caros à comunidade jurídica, no Estado Democrático de Direito, que de certo modo se veem ameaçadas por essas possibilidades de tomada de decisão, recomendação ou realização de inferências sem que a gente tenha uma explicação. E eu acho que, especialmente no campo do Poder Público, existe inclusive uma discussão sobre em que medida essas técnicas de tomada de decisão são compatíveis com o devido processo legal, com a motivação das decisões administrativas, com contraditório, ampla defesa, e assim por diante. Então, quando se fala em explicabilidade e transparência, a gente tem que levar em consideração,

de um lado que muitas vezes nem o próprio autor daquele sistema, nem o cara que programou o algoritmo vai ser capaz de entender porque ele chegou a uma decisão, e de outro lado o fato de que as formas de tomada de decisão algorítmica muitas vezes não são sequer compreensíveis para nossa lógica, a nossa semântica humana, e daí que quando se fala em transparência, me parece também que é necessário superar um pouco uma visão cartesiana, binária de que transparência seria revelar o código fonte, isso daria muito pouco, seja porque o código fonte por si só é passível de auto aprimoração recursiva por meio de *machine learning*, seja porque a gente não consegue entender. Nós leigos não conseguimos entender um código dessa natureza, então por isso é que se tem falado mais em parâmetros de explicabilidade, quanto aos critérios, quanto aos procedimentos e mesmo esse aspecto eu acho que tende a ficar mais desafiador porque a tecnologia tende a evoluir para ficar mais complexa e não mais explicável. Então, sem dúvida, são questões muito difíceis que a gente tem que enfrentar. (E9)

As respostas dos entrevistados vão ao encontro do que Miller (2019) aponta. A ideia de explicabilidade satisfatória da IA, para o autor, tem como ponto de partida a explicabilidade das decisões humanas, mas mesmo essas têm natureza controversa.

O que é uma explicação suficiente e aceitável? Miller (2019) afirma que embora as pesquisas nos campos de ciências sociais (psicologia, filosofia etc) não tenham ainda chegado a um consenso, estão muito mais avançadas e devem servir de base quando se busca uma IA explicável. A conclusão do autor é exatamente o que respondeu o E2 do G1:

[...] talvez o que explicabilidade esteja muito mais no campo de humanas do que talvez até de exatas, de falar o que é que a gente ficaria satisfeito, o que deixaria satisfeito ter como resposta para saber da máquina. Me deixaria satisfeito o fato de eu saber que se eu tivesse uma outra entrada eu teria uma outra decisão? É isso que eu quero? Então eu preciso definir o que eu quero dessa máquina isso é a parte mais complicada da explicabilidade. (E2)

A transparência, por sua vez, está intimamente relacionada com a explicabilidade e seu conteúdo, como colocado por Koivisto (2016), ultrapassa a publicização ou notícia do tema. Em especial em projetos de aplicação de IA ao Judiciário, sua anatomia perpassa a compreensibilidade dos resultados e usos. O cidadão jurisdicionado precisa minimamente entender quando o processo utiliza mecanismos de IA e quais os riscos envolvidos (VALENTE; SUGAI; JUNQUILHO, 2021).

Embora sejam fundamentais a transparência e a explicabilidade, na concepção dos entrevistados, são algo que as ciências sociais devem definir muito mais que a matemática. Isso é, a explicitação do código algorítmico utilizado não basta. E, nesse

sentido, embora não se saiba o que são precisamente explicabilidade e transparência, sabe-se o que não são: não são o mero revelar do código fonte.

A solução para enfrentar a opacidade algorítmica é complexa e, por isso, deve passar pelo debate multidisciplinar que, aliando conhecimentos humanos variados, separe as diversas aplicações da IA em riscos e compreenda a explicação satisfatória e tecnologicamente possível em cada situação.

Um ponto chave, comum entre os respondentes, foi a compreensão de que quando se aplica IA ao Direito, está-se diante de decisões que afetam em alguma medida os direitos dos cidadãos. Procedimentos de caráter público e que, portanto, de algum modo, precisam poder ser questionados e devem ser fundamentados.

4.5.3.2 Opacidade e *Tradeoffs*

A questão da transparência e explicabilidade algorítmica também gera dois *tradeoffs*⁷³ apontados por Pasquale (2015), Frazão e Goettenauer (2021) e Barredo Arrieta et. al. (2020), que suscitaram a criação da subcategoria “*tradeoffs*”.

Pasquale (2015) utiliza o termo caixa-preta, para definir a opacidade das tecnologias mais avançadas de aprendizado de máquina que, protegidas pelo segredo de negócio, não permitem a inteligibilidade dos eventuais erros, o controle e o entendimento pleno das aplicações. Frazão e Goettenauer (2021) descrevem como “dilema jurídico”, o impasse entre a transparência e o regime de proteção da propriedade intelectual. Quando se compreende que a transparência algorítmica vai muito além da divulgação dos modelos matemáticos – mesmo porque a complexidade de criação dos sistemas envolve dados e pessoas – esbarra-se na proteção do segredo de negócios, que inviabiliza sua inteligibilidade.

Essa compreensão foi evidenciada pela E3 do G1:

[...] Boa parte do poder desse universo digital está nas *Big Techs* e o funcionamento é todo em segredo. Então como é que você vai fazer um debate público se um poder está desequilibrado? Então, a transparência e a explicabilidade passam também por ajudar a reequilibrar o poder para pensar a democracia digital. Porque do jeito que está não dá para pensar a democracia digital. Porque é uma briga bem desigual. Em especial a explicabilidade [...] seria interessantíssimo não só a sociedade civil entender e ter autonomia de escolha do que ela quer que seja usado com alguns dados, com os dados delas ou não... se ela quer usar ou não... mas pensando até ao

⁷³ Termo utilizado nesta pesquisa com o sentido de escolha difícil, complexa.

futuro educacional, jovens e crianças que estão hoje imersas nesse futuro digital, você ter a explicabilidade sendo colocada com uma linguagem acessível, você ajuda a desenvolver o que nós chamamos de *digital literacy* ou *artificial intelligence literacy* para jovens e crianças pensando realmente o futuro ... aí você vai estar criando uma estratégia a longo prazo ... e não estar criando uma estratégia dessas a curto prazo de botar um monte de gente para aprender a programar *python* e beleza. Ninguém raciocinou, ninguém refletiu sobre o que aprendeu a programar, só aprendeu a programar e virou quase uma máquina, porque ele fica condicionado a pensar que todos os resultados serão gerados da mesma forma. (E3)

No G2 o E5 também ressaltou o mesmo dilema:

[...] a transparência é desejável e tem muitos que defendem a possibilidade de auditoria dos algoritmos. Perfeito, mas a partir do momento que nós temos algoritmos comerciais, nós temos também a proteção industrial para que nós possamos preservar a livre iniciativa preservar o mercado e tudo mais. É muito complexo o sistema por causa disso. (E5)

E a E8 do G3:

[...] para que haja explicação ou para que a gente tenha uma possibilidade de uma IA explicável/explicabilidade é necessário que o *software* deva ser transparente e auditável. Aí a gente tem uma questão que é fundamental, no que diz respeito a essa transparência e essa possibilidade de auditabilidade do sistema, que é o fato de que muitos daqueles que desenvolvem e aplicam sistemas de IA utilizam argumentos do segredo comercial ou industrial para não permitir o acesso aos dados, à metodologia utilizada para alcançar os resultados previstos ou pretendidos pelo sistema. Então, para que a gente tenha uma IA explicável é necessário que ela seja transparente ainda que não seja totalmente transparente, mas que seja transparente em relação aos dados que foram utilizados, ao percentual de importância que é dada a cada uma dessas categorias de dados para se chegar à uma determinada aplicação da IA. E é fundamental que seja dada a oportunidade às pessoas humanas de reconhecerem que essa é a metodologia ou que esse procedimento na tomada de decisões pela IA seja de fato transparente ou que seja considerado para o titular de dados o direito de reconhecer como a decisão foi tomada pela máquina. (E8)

Outro *tradeoff* destacado pela literatura (Barredo Arrieta et al., 2020) e pelos respondentes da área técnica (G1) é o paradoxo de que em geral, quanto maior a acurácia e precisão dos modelos de redes neurais profundas, maior a dificuldade de explicação e transparência. Isso é, em geral, quanto maior a precisão dos resultados obtidos pelo modelo de IA utilizado, menor a possibilidade de interpretabilidade dos fatores que geraram aquele resultado.

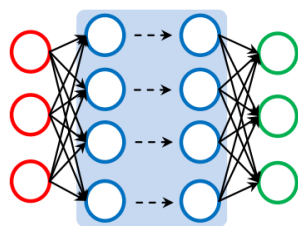
Esse *tradeoff* apareceu em respostas dos entrevistados do G1, como no E1:

[...] Existe uma área chamada *explainable AI* e existe um *tradeoff*. Geralmente os modelos mais assertivos – geralmente (nem sempre) – não são

baseados em uma lógica formal. [...] A área de aprendizado profundo (*deep learning*), ela usa modelos que têm centenas. às vezes milhares de camadas. Nós estamos falando aí de otimizar milhares, milhões, às vezes bilhões de variáveis. Como é que explica isso? Bom, explicar isso do ponto de vista da lógica não se explica. Agora como que você pode explicar isso também? Você pode pontuar assim: ao invés de falar de explicabilidade do modelo de IA, por que nós não falamos de assertividade do modelo de IA? (E1)

Como já se teve oportunidade de abordar nessa pesquisa (ver em especial tópicos 1.1.2 e 3.1.1), os modelos mais utilizados atualmente na IA aplicada ao Direito e que garantem os resultados mais precisos de previsões, são os de aprendizagem profunda. Acontece que, como mostra a figura abaixo, nesses modelos as entradas e saídas são conhecidos, mas a forma como se chegou ao resultado não. Nesse exemplo simples de rede neural profunda pode-se perceber “que as entradas dos algoritmos de aprendizagem profunda são transformadas por meio de várias camadas ocultas e as saídas são derivadas do cálculo das camadas ocultas” (Hao & Zhang, 2016).

Figura 18- Exemplo de funcionamento de algoritmos de aprendizagem profunda



Fonte: HAO; ZHANG, 2016, p. 419

Quando se trata de projetos que envolvem modelos mais complexos de aprendizagem profunda no Poder Judiciário, esses *tradeoffs* apontados pelo marco teórico e pelos respondentes evidenciam-se como entraves ou dificuldades para a concretização dos comandos constitucionais da necessidade de fundamentação das decisões judiciais e, como consequência, do respeito ao devido processo legal. Projetos em aprendizado profundo colocam nas mãos do gestor público executor responsável pelos riscos do programa, maiores desafios, na medida em que em algum momento pode ser necessária a escolha entre um modelo mais opaco e mais eficaz ou menos opaco e menos eficaz de IA (mesmo sem que haja balizas científicas para compreender entender quais critérios usar no momento dessa definição).

4.5.4 Dadaísmo judicial

No contexto da Primeira Guerra mundial (1914-1918), na Suíça, surge o movimento dadaísta. O Dadaísmo, enquanto movimento cultural autocontraditório, de negação da ordem, de reconhecimento do caos e do absurdo, orientou a categoria “dadaísmo judicial”, usada nessa pesquisa para representar o contexto atual de caos provocado pela era do *Big Data* e da Internet das Coisas. O manifesto dadaísta explica a concepção da palavra dadá: “Dadá é tudo, dadá é nada” (TZAR, 1918). Não há lógica no conceito.

No Brasil diz-se que o movimento inspirou a Semana de Arte Moderna, assim como os autores Manuel Bandeira e Mário de Andrade (em especial na obra *Macunaíma*) (Coronato, 2018). Dadá enquanto categoria da pesquisa representa a babel que atualmente é a captura e uso de dados pessoais (sensíveis ou não) na produção de IA para o Direito.

Os entrevistados do G1 mostraram preocupação com a falta de estruturação do dado judicial.

[...] No cenário do Brasil [...] a gente só não está mais avançado por uma questão: a gente está tendo que entregar soluções com o bonde andando [...] mas no mundo ideal, numa situação mais real, eu acho que seria muito importante parar e refletir sobre o dado do judiciário. Se a gente tivesse um dado mais estruturado a gente estaria passos largos à frente [...] e a gente não encontra esse ambiente favorável. O dado do judiciário é um desafio, primeiro, pela qualidade de entrada desse dado. Há uma flexibilidade muito grande, há uma abertura muito grande de protocolação de petições iniciais, então tem vários formatos (até foto de celular). Você encontra de tudo quanto é forma, você encontra o dado ali, então aí eu tenho batido nessa tecla: a gente precisa, à luz da justiça 4.0, dessas novas normatizações que estão saindo do CNJ principalmente. A gente precisa repensar a estruturação do dado do judiciário. [...] um ponto de vista sim estratégico tático. Eu acho que a gente precisaria também pensar não só na aplicação direta de IA mas pensar em um projeto maior aí de estruturação do dado do judiciário. Eu não tenho dúvida que isso ajudaria muito e traria uma economicidade muito grande para todos os processos projetos de IA que estão acontecendo todos os que nós elencamos aí outros que virão. (E1)

Essa preocupação com a estruturação do dado judiciário mostrou-se uma constante. O Poder Judiciário brasileiro e seu cenário atual de abarrotamento processual possui, de fato, um imenso potencial para desenvolvimento de aprendizado de máquina, cuja matéria prima é a existência de uma grande massa de dados. Porém os dados processuais ainda são de difícil acesso a quem pretende, de fora do Judiciário (cidadãos, advogados, pesquisadores e iniciativa privada), desenvolver esses sistemas (BECKER;

TORRES, 2021). E, além disso, os dados judiciais advindo dos processos carecem de sistematização e uniformização, o que se traduz em dificuldade ou obstáculo a qualquer projeto (JUNQUILHO; MAIA FILHO, 2021).

Apesar da existência de regulação protetiva de dados pessoais, torna a estar presente nas respostas dos entrevistados do G1 a preocupação com a responsabilização e falta de rumo quanto ao modo de utilização de dados pessoais processuais, em especial os sensíveis. Como se percebe na resposta do E1:

[...] como que a gente pode trazer tranquilidade para quem vai desenvolver um projeto? Como a gente pode fazer isso? Tem alguns caminhos aí: termo de sigilo, termo de responsabilidade, mas enfim... A gente precisa encaminhar nisso. Os conselhos de ética e não sei eles têm algumas experiências pessoais na área de saúde com isso o judiciário está começando a ter aplicação de IA ele tem dado sensível também talvez eles viram algum modelo aí né [...] **eu não tenho resposta para isso**. Eu só tenho receio de se decidir uma normativa aí com né eu não sei ao pé da letra comentar a LGPD. Todavia, se for uma normativa que seja muito restritiva a gente cai no risco de ir diminuindo, esfriando esse avanço na frente de ciência de dados em dados jurídico. Eu acho que é outro ponto também. Nós falamos do ponto da qualidade do dado, da situação do dado, e é também essa questão aí da sensibilização do dado sensível não é **como que a gente pode trazer uma tranquilidade jurídica para ambas as partes: quem está demandando e desenvolvendo?** (E1, grifo nosso)

Também se nota na resposta do E2

[...] utilização de dados pessoais sensíveis você não tem necessidade nenhuma de utilizá-los, porque você pode simplesmente filtrá-los na hora da própria preparação do *dataset*. Então a gente está falando de dados pessoais não sensíveis. Se você pega já na hora que você vai fazer a *dataset* e você tem que separar o máximo de dados sensíveis jogar fora, para colocar no *dataset* ou não, porque é aí tem uma questão delicada e aí eu não vou entrar muito nisso daí, mas é o seguinte: às vezes, se você tirar os dados sensíveis você está tirando dado de criança, por exemplo.

Entendeu? Então às vezes você acha que você está tirando e fazendo bem, você está tirando e fazendo mal. Então é uma questão delicada para ser discutida em um grupo que tenha conhecimento, que tenha essa capacidade discernimento tanto jurídica, quanto de dados, quanto ética. É uma questão supersensível, não tem uma resposta fácil. A resposta mais fácil é: se é sensível, tira. Você tem certeza disso? Às vezes tirando você está criando problema

[...] a LGPD exige anonimização. A anonimização, ela pode ser gerada com mais eficiência se antes de gerar anonimização você souber quais dados você quer, ao invés de você anonimizar, você simplesmente pega os dados que você quer e não pega os outros (E2)

E na resposta do E3:

[...] As pessoas da sociedade civil não estão a par de tudo o que está acontecendo. Então, antes de você usar dados de pessoas sensíveis ou não para gerar suas pesquisas e isso até mesmo na ciência de dados que as empresas coletam nossos dados o tempo inteiro e ficam usando para

pesquisas que a gente nem conhece, a gente nem sabe. A questão ética precisa estar presente [...] ninguém coloca transparência, ninguém traz a transparência com um fator principal da sua solução então realmente eu acho muito complicado sendo dado sensível ou não. (E3)

Apesar da recente entrada em vigor da LGPD, os procedimentos legais relativos à coleta de dados previstos na legislação, por sua complexidade, têm gerado confusão na área prática e técnica nos desenvolvedores, assim como dúvidas jurídicas a quem aplica a lei (WIMMER, 2021).

No G2 observou-se a mesma inquietação com o modo de coleta desses dados. Em especial destaca-se a resposta da E4

[...] A Resolução 332 no artigo 15 ou em um dos artigos ele fala que os dados utilizados nesse processo de treinamento do sistema de inteligência artificial têm que ser protegidos contra riscos, não de modificação, inclusive de acessos indevidos, acessos não autorizados. Ele diz que o uso desses sistemas de inteligência artificial tem que ser feito levando em consideração as determinações da lei geral de proteção de dados. Então os processos judiciais envolvem por sua natureza dados sensíveis. A gente tem ali dados, por exemplo, direito previdenciário em que a pessoa está pedindo um auxílio-doença, ou seja, ela está incapacitada para o trabalho em razão, digamos, de uma patologia. Aquele dado de saúde que estava naquele processo ele é um dado sensível. Eu já peguei processos em que um dos pedidos, uma das patologias, era o fato da pessoa ter HIV e ela não queria que na audiência eu perguntasse sobre o HIV porque as testemunhas iam ouvir e saber que ela tem. Não queria que ela soubesse, que as testemunhas soubessem. Então isso é um dado extremamente sensível e é sensível com base na lei, na legislação porque seus dados [...] esses dados utilizados para o treinamento de um sistema... se está falando, por exemplo, de um sistema que vai utilizar dados previdenciários, acaba que ele vai pegar muitos dados sobre saúde, porque vai falar sobre direito previdenciário com relação a alguns benefícios, envolve necessariamente falar sobre laudos periciais, sobre laudos médicos, então é um cuidado ainda que tem que ser dobrado. Porque não são só dados de pessoas como são dados sensíveis. Agora é por isso que a gente tem que ter cuidado com: primeiro ponto quem que está treinando [...]? É o poder público ou é a iniciativa privada? Se a gente deixar a iniciativa privada fazer parcerias com o poder público, porque a iniciativa privada tem muito mais expertise do que o poder público nessa área, a gente tem uma infinidade de *startups* que são responsáveis pelo avanço da tecnologia, porque a gente deixar que o poder público faça então, inclusive o judiciário, em especial judiciário, faça parcerias com a iniciativa privada ou quanto do acesso ao pessoal da iniciativa privada vai poder ter e como que eles vão, depois, como vai ser feito o descarte desses dados, qual a finalidade que eles vão dar para esses dados depois daquilo? Isso tem que estar muito bem estruturado, inclusive na parceria, no contrato, seja qual for a forma com que vão chamar esse negócio jurídico que vai ser feito da iniciativa privada com o poder público (E4).

[...] é necessária a independência da fonte de financiamento dessa plataforma [...] ou seja, é pra gente **diminuir essa chamada assimetria de informação** e tornar o litigante eventual um pouco mais próximo um pouco mais, porque não vai conseguir ficar igualado, mas um pouco mais próximo do litigante habitual, seja uma fase de diagnóstico daquele problema, seja na fase da negociação facilitada pela tecnologia, seja inclusive na fase em que são feitas

efetivamente propostas de acordo com o uso da inteligência artificial. O desenvolvedor daquele sistema tem que ser um desenvolvedor independente, interesse naquele conflito e que não seja parte naquele conflito. Essa me parece ser uma forma que é possível a gente admitir a iniciativa privada em parceria com o poder público. Parceria estou dizendo assim lato sensu, né, pode ser qualquer forma de negócio jurídico, mas que a gente consegue ainda assim resguardar uma proteção tanto para o resultado final, quanto para os próprios dados das partes que estão sendo utilizados ali. (E4, grifo nosso)

A E4 cita ainda a necessidade de garantir a paridade de armas processuais ao litigante eventual. E, nesse sentido, que as informações obtidas pelas partes em conflito estejam disponíveis de forma equânime para os cidadãos e empresas⁷⁴.

É nítido que o uso de dados pessoais extraídos de processos judiciais, para aplicação de IA ao Direito, tem suas especificidades. O dado processual não é estruturado, é público (em regra), mas de difícil acesso, além do que, eventualmente, podem ser capturados dados sensíveis, etc. Talvez por essa razão, mesmo existindo lei específica no Brasil para disciplinar o direito constitucional à proteção de dados, restam nas respostas dos entrevistados mais dúvidas que certezas sobre o tema.

É possível coletar dados sensíveis? Proibir uso de dados sensíveis em projetos de IA para o Judiciário pode excluir dos projetos os grupos que se pretende proteger na privacidade? Como definir a base de treinamento adequadamente representativa? Como realizar a anonimização em determinados projetos que usam dados processuais desestruturados? Como viabilizar a abertura dos dados judiciais ao acesso público? Quem e como controlar a utilização ética dos dados pessoais processuais?

Há, inclusive, certa angústia dos respondentes em saber que os dados pessoais estão sendo utilizados com certo descontrole, com a dificuldade de *accountability* e com a ausência de segurança nos rumos da responsabilidade do desenvolvedor. A categoria “dadaísmo” foi usada aqui, portanto, para representar o contexto atual de caos provocado pela enxurrada de dados pessoais processuais que são úteis aos avanços da IA no Judiciário brasileiro, mas que, por outro lado, mesmo os especialistas que entendem do tema se sentem confusos em definir a melhor forma de controle e limitação.

O “dadaísmo judicial” representa, então, essa desordem. “DADÁ, DADÁ DADÁ, alarido de dores crispadas, entrelaçamento dos contrários e de todas as contradições, dos grotescos, das inconseqüências: A VIDA” (TZARA, 1918).

⁷⁴ Aqui se recomenda rememorar a questão ética trabalhada no tópico 3.1.2, acerca da paridade de armas e da jurimetria.

Cabe destacar que uma única entrevistada do G3 destoou das demais respostas:

Existe já para esse assunto específico de dados pessoais e dados pessoais sensíveis uma legislação que é bastante protetiva no que diz respeito aos interesses patrimoniais iniciais do titular de dados, então eu vejo que essa regulação por si só já seria suficiente para dar algum tipo de limite um caminho impossível é para uso da inteligência artificial ético não é do nosso país. (E8)

Entendeu, assim, que a LGPD é legislação suficiente para limitar os dados pessoais utilizados para produção de IA.

4.5.5 Vieses como “limonada pronta”

Os algoritmos são denominados por Cathy O’Neil (2016) de máquinas de destruição matemática, pois, conforme a autora cita em exemplos, podem perpetuar desigualdades. Os modelos matemáticos não são neutros, ao contrário, carregam em si os preconceitos humanos advindos dos dados passados e da classificação humana dos desenvolvedores.

Mas como os entrevistados encaram os vieses nos resultados das aplicações de IA ao Direito?

[...] Eu vejo uma maneira muito positiva a questão dos vieses. Eu olho o lado positivo. Quer dizer o seguinte: às vezes as pessoas só falam isso de uma maneira assim de ser um limão. Para mim é uma **limonada pronta**, você sabe por quê? [...] Os modelos de IA aprendem com os dados históricos. Se eventualmente você tem uma IA que ela apresenta vieses, e aí eu estou entendendo os vieses que você está falando aí, por exemplo, de ter um direcionamento por raça, cor, preferência sexual ou enfim. Nada mais é do que ela aprendeu com o dado. E aí? Ser humano que fez a IA. Ser humano que está usando. Você vai fazer o que com isso? Eu acho que isso é uma autocrítica que eu acho muito interessante... está aí, Brother, mas você vai fazer o que? Vai continuar fazendo assim ou você vai mudar? Ah... eu quero mudar... então beleza, vamos fazer um trabalho de análise exploratória de dado? Vamos treinar uma outra IA? Bora! Somos nós. Somos nós seres humanos que fazemos a sociedade. Não é a IA. (E1, grifo nosso)

A partir da resposta do E1 do G1, criou-se, então, a categoria “vieses como limonada pronta”, que representa o fato de que a maioria dos respondentes entende que o uso de algoritmos no Judiciário pode desvelar a existência de vieses discriminatórios de quem está envolvido no processo decisório e, inclusive, ajudar a corrigi-los (fazer do limão – o histórico de enviesamento discriminatório – uma limonada).

[...] Mas isso é uma boa solução para tentar amenizar a questão do viés. Você saber exatamente o que está sendo levado em consideração na hora de fazer o ranqueamento, a classificação, agrupamento. Se você consegue fazer isso você consegue aliviar bastante esse problema de viés.

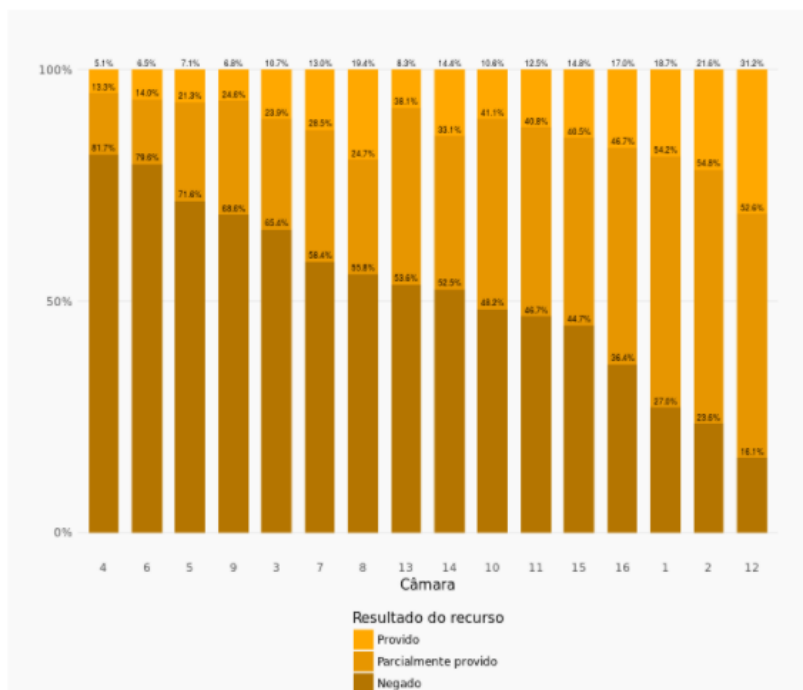
[...] Agora tem um outro problema de viés que é um viés do magistrado [...] Se eventualmente tiver um magistrado que tenha viés discriminatório, a máquina vai poder reproduzir isso. [...] porque a máquina está espalhando o erro [...] a máquina aprende com as decisões daquele magistrado específico [...] o cientista sênior de inteligência artificial ele não tem uma capacidade de julgar o juiz então o que é que a máquina faz? A máquina ela vai tentar fazer aquilo que aquele juiz fez ontem. É isso que a máquina vai tentar fazer. E se aquele juiz tem um viés discriminatório não deveria ter. Mas não é a máquina que vai corrigir, é um problema que existe. Se não tivesse a máquina, continuaria tendo esse problema, a máquina não gerou esse problema, é porque existe em outras áreas uma questão muito... é que a máquina... é reproduzir o viés discriminatório porque o viés... primeiro pode existir os dados, mas quando a gente está falando do direito, a gente chega num problema que é o seguinte, é diferente de outras áreas que podem, por exemplo o *Facebook* pode chegar a falar o seguinte: olha esse post que você postou isso aqui é discriminatório eu vou excluir esse post. O *Facebook* faz isso né ...não vou entrar no mérito se você pode ou não pode mas ele faz isso... agora o cientista sênior de computação de inteligência artificial, ele não pode chegar e falar: essa decisão do juiz está errada acho melhor não acho melhor ele não fazer isso aqui porque isso aqui na minha opinião é discriminatório... você pode fazer isso então, o que ele pode fazer ele pode evitar que aqueles conceitos daquele juiz seja imposto outro juiz. (E2)

Sunstein (2019) e Kahneman; Sibony; Sunstein (2021) também apontam que ao utilizar dados passados, os algoritmos podem revelar nos resultados, os vieses historicamente presentes nas situações em que é empregado⁷⁵. Nesse sentido, veja-se o exemplo do recente estudo realizado pela Associação Brasileira de Jurimetria (ABJ), que demonstrou como algumas Câmaras do Tribunal de Justiça de São Paulo (TJSP) tendem a revisar ou não os recursos criminais:

Figura 19– Reformas de decisões nas câmaras criminais do TJSP

⁷⁵ Cabe ressaltar também que a maioria dos respondentes citou como exemplo de viés no Direito o caso do software *Compas*, já descrito aqui em capítulo anterior.

Reformas de decisão nas câmaras criminais



Fonte: <https://abj.org.br/cases/camaras-2/>.

A pesquisa jurimétrica da ABJ tal qual o Relatório produzido pela ONG ProPublica sobre o Compas, rompe com o mito da imparcialidade da justiça, ao apontar exatamente a ampla disparidade entre o resultado de recursos a depender da Câmara em que ele é julgado.

As análises preditivas realizadas pela IA, nesse sentido, além de reduzir ruídos, podem indicar, por exemplo, os grupos minoritários mais atingidos pelas decisões. Mas precisam, para tanto, de vontade humana de correção da situação. Isso é, um estudo como o da ABJ que aponte as tendências decisórias do TJSP sem ações que alterem o atual cenário, não corrigem os vieses nem diminuem os ruídos.

O fato de que a tecnologia não é neutra, pois pode levar a resultados discriminatórios que decorrem da qualidade da fonte de dados ou dos critérios humanos para o desenvolvimento das aplicações de IA, esteve presente em todas as respostas, corroborando a discussão de que os modelos de redes neurais são de difícil inteligibilidade (VILONE & LONGO, 2020). Destacam-se duas delas a seguir:

[...] A gente tem que sempre lembrar que a máquina reflete quem nós somos enquanto pessoas. O velho argumento de que a tecnologia é neutra, na realidade nenhuma tecnologia é neutra, nem a tecnologia que era aplicada na revolução industrial era uma tecnologia neutra. Então, um sistema de

inteligência artificial é tão neutro quanto for a pessoa que vai desenvolver e aplicar o sistema, ou seja, é impossível ter neutralidade quando a gente está diante de tecnologia. E no caso da inteligência artificial, a gente tem dois problemas: o fato de que não existe uma multiplicidade, a diversidade de pessoas trabalhando no desenvolvimento na programação do sistema de IA... a gente diz brincando, mas verdadeiramente, que o programador é o homem branco, heterossexual, classe média ou classe alta que teve acesso a uma educação de alta qualidade. Isso é verdade, sem dúvida nenhuma. Então o mundo desse sujeito é o mundo que é refletido na programação. Quanto mais diversa for, quanto mais diversos forem os grupos que trabalham no desenvolvimento de sistemas, mais diversos serão os resultados esperados [...] Isso é uma coisa. Segundo problema é o fato de que as pessoas são reconhecidamente preconceituosas no sentido de que elas têm pré concepções, de ter determinadas realidades de vida. Então necessariamente esse viés vai aparecer, porque se essa inteligência artificial trabalha com dados, esses dados reproduzem a realidade. Então, é possível correção ao longo do tempo e esse tipo de fiscalização (e tem que ser feita essa fiscalização), ela vai representar justamente um dos princípios aplicáveis a IA que é o princípio da *accountability*, que é o princípio da prestação de contas. O sistema transparente, a possibilidade de auditar e a possibilidade de prestar contas em relação aos resultados alcançados. (E8)

[...] De um lado a gente sabe que bases de dados de treinamentos que sejam inadequados e sejam insuficientemente representativos, tendem a fazer com que a máquina aprenda a reproduzir esses vieses, já existentes. [...] A gente tem uma outra questão associada, não às bases de dados, mas às próprias técnicas de *machine learning* que é o que faz com que as máquinas acabem por *Proxy* aprendendo certos comportamentos que são discriminatórios [...] E por fim, a gente tem uma situação que é mais difícil ainda [...] ainda que a gente consiga dizer para a máquina que não deve considerar os fatores raça, religião, sexo, ainda que a gente consiga tirar tecnicamente esses vieses, a gente ainda vai se deparar com situação em que decisões algorítmicas vão afetar de maneira desproporcional determinados grupos étnicos ou sociais, ainda que eles não sejam discriminatórios por si, eles vão reforçar determinados tipos de vieses já existentes ainda que sejam vieses matematicamente explicáveis. [...] E isso nos conduz a um debate muito mais filosófico que é: reconhecendo que discriminar faz parte, é elemento essencial das políticas públicas, da vida em sociedade, segregar grupos, para tratá-los de maneira desigual, quais são as discriminações efetivamente ilícitas, abusivas, que não deveriam ser admitidas, e como exercer controle sobre isso, como visualizar que isso de fato acontece, como corrigir? E aí isso nos leva para uma outra discussão de governança, de transparência de *oversight*. Acho que é um tema tremendamente complicado, acho que a gente está começando a buscar compreender um pouco melhor e me parece que não tem uma bala de prata. Não se trata apenas de cuidar das bases de dados de treinamentos porque depois tem outros problemas que podem surgir ainda que as bases de dados estejam bem equilibradas, bem diversas. Então esse é um problema da nossa geração. (E9)

Como afirma Birhane (2021, p. 176)

[...] a demanda por “resolver” problemas sociais com a tecnologia é promissora e é exatamente daí que surgem os problemas. Problemas culturais, morais e políticos complexos, inerentemente imbuídos na história e no contexto, são reduzidos a problemas que seriam de tecnologia.

Ora, a IA não é capaz de solucionar com neutralidade os vieses, primeiro porque a IA não é neutra, segundo porque os vieses são problemas muito mais complexos, que advém de fatores culturais. No Brasil citem-se como exemplo o passado/presente escravocrata, a colonização e o ranço de conservadorismos da política ditatorial.

Como afirmou a E3 do G1

[...] quando a gente vê o debate do viés acontecendo, até usando essa palavra muitas pessoas pecam porque elas colocam como algo solucionável pela tecnologia. Se tivesse solução a partir da perspectiva tecnológica a gente não precisaria estar aqui nesse debate, você não precisaria estar fazendo o seu trabalho. Isso está se naturalizando porque o mercado precisa de argumentos para constar e não quebrar o negócio dele. Então, a partir do momento que o seu negócio é atrelado ao algoritmo, se o algoritmo é parte do negócio, além de prejudicar a empresa em relação à transparência e tudo mais. Por isso que a gente tem todos os problemas para entender os algoritmos dessas grandes *big techs*. [...] se você for olhar, já está na boca de todo mundo viés, conselho de ética, não sei o quê... e gente, isso não está resolvendo o problema, na verdade isso são coisas muito inúteis, porque ela depende de todos os agentes que estão na sociedade. [...] Não vai sair de uma pessoa, de um pesquisador, são estratégias. **É uma estratégia coletiva. Precisa ter essa noção de que é uma estratégia coletiva, de que é um problema da sociedade, que é um problema na sociedade que não é com perspectiva tecnológica** e quando a gente sai um pouco desse olhar da tecnologia que é um exercício que eu venho fazendo há bastante tempo, você entende a partir de outra perspectiva, você compreende, por exemplo, que em uma lei como a LGPD precisam ser reescrita algumas coisas ...[...] **Então o viés não é um viés que está só sendo reproduzido pelo algoritmo, é um viés de padrões. São padrões de uma sociedade, são padrões de comportamento.** (E3)

Se os vieses discriminatórios da IA refletem os vieses humanos, um caminho tecnológico inicial para sua possível correção e/ou para evitá-los é a exigência do uso de bases representativas e de equipes desenvolvidoras diversas e plurais. A realização de auditorias constantes nos algoritmos é ação fundamental também para denunciar a invisibilidade de determinados grupos, o racismo, a discriminação e até identificar micro agressões cotidianas geradas pelos vieses (SILVA, 2021).

Os vieses podem ser de fato “limonada pronta”, mas não porque a IA é capaz por si enquanto tecnologia de corrigir discriminações. Como apontou a E3, há que se pensar que a IA não é neutra, que ela pode desvelar e expor preconceitos de fato, mas que a opressão a grupos minoritários é eminentemente humana e sua correção é uma estratégia coletiva complexa que depende de vontade política, de mudança de aspectos culturais e de muitos outros fatores sociais.

4.5.6 “E agora, José?”

“E agora, José?”, poema em que Carlos Drummond de Andrade⁷⁶ questiona o destino da vida e para onde marchar, seguir a diante, inspirou a categorização do que os

⁷⁶ No poema “José” Drummond questiona os destinos da vida de José. A categoria aqui inspira-se no poema para avaliar as respostas dos entrevistados sobre sugestões para controle da IA aplicada ao Direito.

José

E agora, José?
 A festa acabou,
 a luz apagou,
 o povo sumiu,
 a noite esfriou,
 e agora, José?
 e agora, você?
 você que é sem nome,
 que zomba dos outros,
 você que faz versos,
 que ama, protesta?
 e agora, José?
 Está sem mulher,
 está sem discurso,
 está sem carinho,
 já não pode beber,
 já não pode fumar,
 cuspir já não pode,
 a noite esfriou,
 o dia não veio,
 o bonde não veio,
 o riso não veio,
 não veio a utopia
 e tudo acabou
 e tudo fugiu
 e tudo mofou,
 e agora, José?

E agora, José?
 Sua doce palavra,
 seu instante de febre,
 sua gula e jejum,
 sua biblioteca,
 sua lavra de ouro,
 seu terno de vidro,
 sua incoerência,
 seu ódio - e agora?

Com a chave na mão
 quer abrir a porta,
 não existe porta;
 quer morrer no mar,
 mas o mar secou;
 quer ir para Minas,
 Minas não há mais.
 José, e agora?

Se você gritasse,
 se você gemesse,
 se você tocasse
 a valsa vienense,

respondentes apontaram como sugestões e caminhos futuros para melhorar o controle ético da IA aplicada ao Direito.

[...] Acho que a gente está numa fase que a gente, em primeiro lugar, está lidando com um **desafio jurídico**, no sentido de que as nossas leis não estão desenhadas nem calibradas ainda para lidar com esses fenômenos. [...] ainda, nós estamos patinando quando se fala de padrões éticos. Me parece que a gente, está muito pouco maduro e recentemente eu tive a curiosidade de fazer uma lista de tentar comparar esses documentos e os vocabulário é muito comum, muito recorrente, mas quando a gente começa a espremer os conceitos, a gente não consegue extrair dali indicadores concretos. Então, trabalhos muito interessantes na comunidade técnica, do que é um design como introjetar parâmetros éticos em sistemas autônomos, mas ao mesmo tempo uma dúvida sobre quais parâmetros éticos seriam esses, e eu acho que quando a gente olha para o dilema do bonde que é sempre citado, e a gente vê como as respostas éticas são dependentes também de cultura, de religião, de contexto, fica evidente a dificuldade de lidar com essa nova realidade, em que decisões serão tomadas ou sugeridas com base em parâmetros que a gente não consegue visualizar. [...] mas de fato desdobrar o que isso significa também nas diferentes áreas de aplicação de IA é conhecendo **que há riscos distintos nessas aplicações é que vai ser nossa grande tarefa daqui para frente**. (E9).

Os respondentes reconheceram a complexidade e o imenso desafio que é a aplicação ética da IA. Algumas ideias de controle comuns nas respostas foram: a necessidade de exigir-se nos projetos de IA relatórios de impacto e a criação de guias de boas práticas para programadores e desenvolvedores envolvidos nos projetos. Alguns respondentes também sugerem separar as aplicações da IA em riscos, tal qual a proposta europeia de regulação indica. No caso do Direito, por exemplo, seria fundamental avaliar áreas em que se lida com hipossuficientes ou bens jurídicos importantes, como a

se você dormisse,
se você cansasse,
se você morresse...
Mas você não morre,
você é duro, José!

Sozinho no escuro
qual bicho-do-mato,
sem teogonia,
sem parede nua
para se encostar,
sem cavalo preto
que fuja a galope,
você marcha, José!
José, para onde?

Disponível em: <https://www.letas.mus.br/carlos-drummond-de-andrade/353799/>. Acesso em: 09 jan. 2022.

liberdade de ir e vir, e avaliar se os riscos impediriam o desenvolvimento desses sistemas.

Em matéria de risco, os respondentes foram igualmente unânimes em considerar que não pode haver decisões autônomas realizadas somente pela IA, sem a supervisão humana, em nenhuma área do Direito. Isso porque o Direito é ciência complexa, que lida com emoções e dores das pessoas e para produzir uma decisão é preciso sentir. O elemento humano é fundamental para analisar provas, analisar sentimentos, reações, perceber contextos⁷⁷.

[...] A atividade humana, ainda que seja essa atividade de julgar, envolve intuições, envolve emoções e muito do conhecimento da espécie humana. Inclusive o conhecimento jurídico, ele é intuitivo. Por mais que digam que existe uma fundamentação judicial que o sujeito estudou anos para chegar naquela conclusão, existe uma parcela daquele conhecimento que é intuitivo. A gente vê nas audiências como que às vezes a forma como que a testemunha dá um depoimento e você percebe que aquilo faz você mudar completamente a impressão que você tinha dos fatos e do depoente, o quão importante é o depoimento do depoente para uma conclusão final. Então muito do nosso conhecimento me parece ser intuitivo e aí para se aproximar disso como é que a máquina vai lidar com essas emoções com essa intuição? Como é que a gente vai inserir isso num sistema de inteligência artificial? E será que é possível a gente criar um sistema de inteligência artificial emocionalmente neutro? (E4)

Assim como Gillis (2021) e Frazão (2021a), os entrevistados deixaram claro que quando as decisões algorítmicas envolvem pontos humanos e sociais, como são as decisões judiciais, as mesmas “necessitam ser complementadas por narrativas e outras metodologias qualitativas para que se possa construir uma decisão correta e que atenda minimamente aos pressupostos da explicabilidade” (FRAZÃO, 2021a). Isso é, decisões judiciais precisam passar, no mínimo, por algum tipo de revisão humana.

Também o *design* ético foi citado pela maioria como forma de controle prévio.

[...] é possível se pensar num *design* que seja eticamente adequado para o desenvolvimento de inteligência artificial. Então, se for possível essa **limitação à priori** não é ex post, posteriormente, a gente sabe que sim né... que através da revisão e implementação das modificações nos sistemas, a priori, se for possível identificar a possibilidade de inclusão no próprio *design* da programação essas limitações éticas seria é sensacional... então, pensando nos princípios éticos mais comumente referenciados, é a equidade, a prestação de contas, a transparência e a sustentabilidade. Esses 4 princípios

⁷⁷ Se a audiência virtual já recebe críticas exatamente pelo fato de expor a parte e impessoalizar o processo, é de se pensar quantas críticas também caberiam para julgamentos realizados exclusivamente pela IA. Texto interessante que critica as audiências por videoconferência no âmbito do direito penal está disponível em: <https://www.conjur.com.br/2020-mai-31/opiniao-problemas-audiencia-virtual>, Acesso em 105 dez. 2021.

podem antecipadamente ser aplicados no sistema, por exemplo, com a confecção de relatórios de impacto, tanto na proteção de dados quanto na implementação de programas de sistemas de IA. A necessidade ou obrigatoriedade, e aí eu acho que nem é mais uma questão ética do dever ser, mas eu acho que deveria ser uma imposição, a necessidade de se construir guias de boas práticas para serem utilizados pelos programadores e aplicadores de sistemas de IA e esses guias de boas práticas seriam limitadores do próprio do próprio desenvolvimento ou da própria aplicação é pode ser que isso seja um desejo um pouco fora da curva, mas eu acho que se a tecnologia é moldada por humanos eu acho que os humanos têm a criatividade suficiente para pensar nessas alternativas de limitação eventual dos vieses e de resultados violadores de direitos fundamentais (E9).

Tais elementos presentes nas entrevistas vão ao encontro do que propõe Dignum (2018), que estabelece como fatores para o desenvolvimento de uma IA responsável a inserção de abordagens de *design* com valores éticos nos produtos da IA. Além disso, sugeriu-se que cada projeto tenha grupos multidisciplinares para avaliação constante dos produtos da IA.

A principal dificuldade e desafio citado foi como trazer para a prática as questões éticas e principiológicas definidas.

[...] eu queria dizer que educação e regulação é você ter claro boas práticas. Sempre é fundamental e eu até clamo que na educação esses métodos, essas boas práticas virem uma disciplina, para além da ética, porque a ética e os princípios... é preciso ter uma aplicabilidade dela, não só na lei, mas também na consecução, na formulação, no desenvolvimento de projetos (E11).

Assim como Floridi et al. (2018), os participantes reconheceram que o debate acerca do controle ético não está maduro. As sugestões estão no campo de compreensão de que diante do *boom* da IA, é fundamental um código de ética, que atribua responsabilidades, estabeleça práticas e esclareça os riscos de seus diversos usos (BODDINGTON, 2017).

5 OUTPUT: TRIANGULAÇÃO DOS RESULTADOS

O objetivo deste capítulo final é realizar a denominada “avaliação por triangulação de métodos”, a qual busca superar a oposição entre os métodos qualitativo e quantitativo procurando, nesse sentido, as melhores contribuições da interdisciplinaridade. “Do ponto de vista da utilidade, uma boa avaliação visa a reduzir incertezas, a melhorar a efetividade das ações e a propiciar a tomada de decisões relevantes” (MINAYO; ASSIS; SOUZA, 2005. p. 19).

Pretende-se, nesse momento, confrontar os resultados dos parâmetros da literatura, da regulação e da percepção de atores, para, então, compreender os fatores e indicadores mínimos de controle ético da IA aplicada ao Direito. Com base na triangulação dos resultados, almeja-se entender e avaliar o fenômeno pesquisado de modo dialógico, produzindo diagnóstico do que foi discutido na pesquisa a partir das diversas perspectivas coletadas nos capítulos anteriores.

Não se pode deixar de reconhecer, como Susskind (2019), Walker (2021) e Surden (2014), as benesses da IA para o Direito, em especial para uma justiça morosa como é a brasileira (um rápido olhar pelo relatório anual da Justiça em Números do CNJ pode confirmar essa afirmação⁷⁸). Não sem razão, mais de 47 tribunais no Brasil desenvolvem atualmente algum tipo de IA (ver Quadro 3) o que, direta ou indiretamente, impacta a vida de milhares de cidadãos jurisdicionados do país.

Entretanto, muitos são os dilemas éticos que surgem (apontados em especial no capítulo 3.1.2 dessa tese) e se colocam como desafio para o equilíbrio entre o estímulo e promoção dessa tecnologia, sem deixar de lado sua necessária limitação.

Embora se espere que isso um dia possa ocorrer (DIGNUM, 2019), atualmente não se consideram os modelos de IA como agentes morais e justamente por não serem compreendidos como agentes éticos em si, é que se torna fundamental avaliar quais as preocupações e precauções éticas que devem ser tomadas quando de sua produção e implementação pelo agente moral humano (Coeckelbergh, 2020). Já que a IA traz tantos impactos e modificações na cultura, na economia e, em especial, para o debate dessa pesquisa, na jurisdição, apresentar o debate da ética no, pelo e para o *design* é realizar o

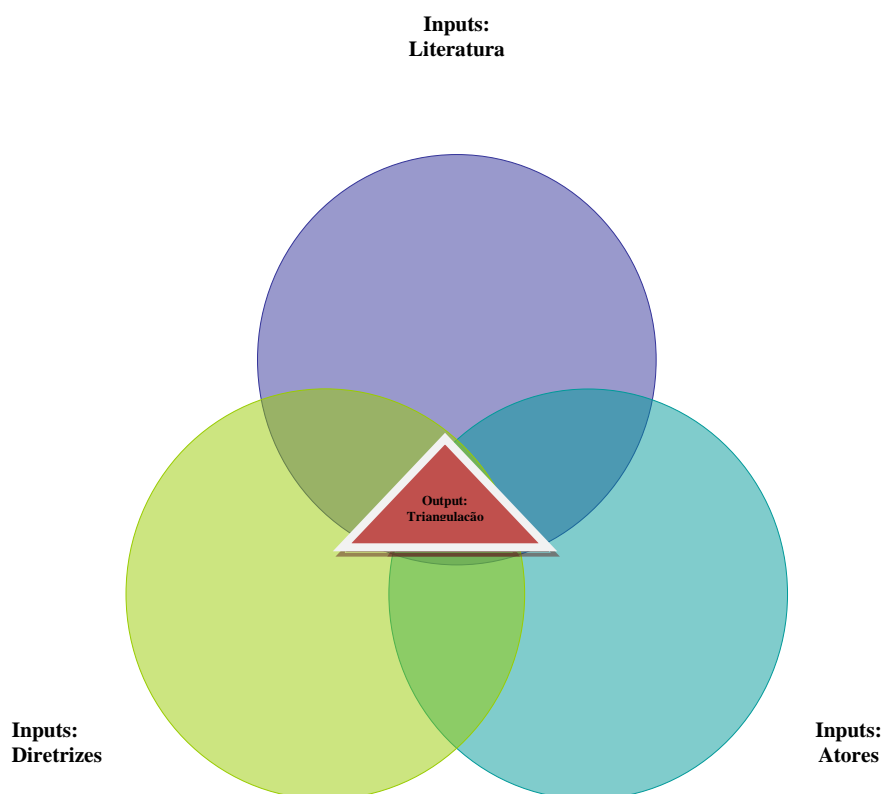
⁷⁸ Ressalte-se aqui o conceito de “giro de acervo” (tempo que levariam os juízes brasileiros para julgar todas as causas existentes, se mais nenhum processo entrasse na justiça), que atualmente é de mais de dois anos (CNJ, 2021).

esforço de compreender qual o papel do humano e dos dados do passado na construção dos modelos atuais da IA.

E qual seria o conteúdo mínimo para o *design* ético da IA quando aplicada ao Direito? Para Walker (2021, p. 148) a IA ética ainda é encarada como “uma cortina de fumaça”.

De modo semelhante ao processo de aprendizado de máquina, esse capítulo pretende descortinar essa questão unindo as conclusões trazidas pelos capítulos anteriores, realizando, então, a avaliação interpretativa dos *inputs* obtidos nas entrevistas, no quadro de referência teórica sobre o assunto e no material regulatório levantado.

Figura 20– Representação da triangulação dos resultados



A análise bibliométrica realizada no Capítulo 2 trouxe variáveis sobre quais os princípios éticos comuns nos textos; quais os problemas trazidos pela aplicação de IA ao Direito aparecem comumente; quais os tópicos do Direito mais citados e as regras mais frequentes.

Os princípios éticos que mais apareceram no conjunto de textos avaliados foram “transparência”, “prestação de contas”, “equidade”, “responsabilidade”, “sustentabilidade” e “explicabilidade”. Aqui se percebeu que há consenso na literatura recente sobre os princípios aplicáveis à IA. Essas variáveis também estavam presentes nas diretrizes avaliadas no Capítulo 3. Isso é, a congruência também está presente nas diretrizes analisadas, pois os princípios mais citados pela literatura estão descritos tanto na Resolução n. 332/CNJ, quanto na CEEUIA SJ, além de também estarem positivados na regra mais citada pelos artigos (o GDPR, marco regulatório que mais apareceu nos textos, apesar de tratar sobre proteção de dados pessoais).

Notou-se, entretanto, que apesar da equivalência entre quais são os princípios mínimos, já que todos os princípios apontados pela literatura estão contidos nas diretrizes, os atores entrevistados no Capítulo 4 apontaram a insuficiência regulatória sobre o tema da ética em IA. Alguns demonstraram, inclusive, a insegurança em trabalhar no desenvolvimento da IA para o Direito, pela falta de regulação que atribua responsabilidades.

Em artigo recente, Floridi (2021, p. 4, tradução nossa) constata ser o fim da era de autorregulação da indústria da IA. Segundo o autor, “a autorregulação precisa ser substituída pela lei; quanto antes melhor. *Dura lex, sed lex digitalis* é a razão pela qual a UE está na vanguarda do debate sobre a governança digital”.

As dificuldades em mitigar os riscos da IA não podem transformar o chamado verão que garantiu as conquistas e progressos dessa tecnologia em nova fase inverno com limitação excessiva. Ou seja, as medidas de controle ético devem ser avanços, que façam o sistema funcionar de forma mais célere, e não entraves técnicos e burocráticos que podem criar novos gargalos.

Floridi (2021) parte do contexto europeu para afirmar que, para além de princípios e diretrizes, que já existem ao redor do mundo, a criação de regras e de sanções tem se mostrado fundamental. Tomando-se por base a contatação do autor e a partir da triangulação dos resultados obtidos nessa pesquisa, pode-se dizer que para o

contexto brasileiro, a autorregulação não só da indústria, mas também a setorial pública e que os princípios que atualmente existem, têm se mostrado insuficientes.

Isso é, embora a atual diretriz setorial pública brasileira (Resolução n. 332/CNJ) e a LGPD prevejam todos os princípios colocados pela literatura, os mesmos não se mostram suficientes para garantir e nortear a atuação dos especialistas e estabelecer o necessário para um *design* ético seguro. Desvincular as decisões das máquinas dos vieses existentes é uma tarefa que vai exigir esforço hercúleo de todos os envolvidos

Os princípios que mais apareceram na literatura, por sua vez, estavam relacionados aos principais problemas da aplicação de IA ou pontos de inquietude trabalhados nos textos. As palavras mais frequentes nesse sentido foram “viés”, “privacidade”, “dados”, “caixa-preta”, “preditiva”/“predição”.

Os maiores riscos que as aplicações de IA podem apresentar, segundo a produção científica mais recente, são a opacidade algorítmica, a invasão à privacidade para utilização de dados e os vieses discriminatórios. Essas preocupações são apontadas pelas diretrizes avaliadas.

A Resolução n. 332/CNJ atenta-se no Capítulo III à não discriminação e no Capítulo IV à transparência, além de tomar por base a consonância com a LGPD.

Os cinco princípios norteadores da CEEUIA SJ (respeito aos direitos fundamentais, não discriminação; qualidade e segurança; transparência, imparcialidade e equidade; controle do usuário) também estão intimamente relacionados aos riscos principais citados.

Os atores entrevistados, por sua vez, apontaram e reconheceram os mesmos riscos da IA, mas tiveram dificuldade em atribuir sentido ou enxergar saídas para mitigação dos mesmos. Alguns, como se afirmou, demonstraram certo receio em desenvolver aplicações de IA para o Direito sem entender precisamente como estão sendo controlados, mitigados ou limitados esses riscos que podem advir.

Mais uma vez, a triangulação demonstra que o debate e a prática estão alinhados. Porém, reforça o que destacou Juliano Maranhão⁷⁹ recentemente: que o cenário da tecnologia de IA carece de uma legislação base que estabeleça requisitos mínimos para o desenvolvimento seguro de qualquer projeto e que garanta a responsabilização responsável, isso é, sem excessos. É o que o autor chama de

⁷⁹ A palestra completa do autor sobre a questão regulatória em IA foi dada para a Cátedra Oscar Sala no Instituto de Estudos Avançados da USP (Universidade de São Paulo) e está disponível em: <https://www.youtube.com/watch?v=nQ6OJNsr-II>. Acesso em: 07 dez. 2021.

“autorregulação regulada”, existência de uma legislação geral pelo Poder Público, que trace certificações mínimas vinculantes de governança para IA e posterior regulação dos órgãos de cada setor que traga as especificidades da aplicação de IA em cada área, realize uma abordagem baseada em riscos, assim como estabeleça formas ou órgãos internos de controle.

Além disso, também foram termos recorrentes na literatura “juízo”, “juiz” e os tópicos “criminal”, “consumidor”, “civil”, “constitucional”, “governo” e “administração”. E, nesse sentido, tanto as diretrizes, quanto as respostas às entrevistas, demonstraram especial apreensão com as decisões autônomas e a aplicação de IA na área penal.

Aqui cabe destacar que, embora esteja presente toda essa preocupação, o art. 23 da Resolução n. 332/CNJ é impreciso ao tratar do tema, afirmando que “a utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada”. O dispositivo estaria prevendo aqui uma permissão sem fomento? Não há clareza sobre a questão.

Nas entrevistas também foi unânime a compreensão dos atores de que a IA não pode gerar decisões autônomas no Direito e que deve sempre haver revisão humana que realize o reexame e chancele as predições realizadas pela aprendizagem de máquina. Afinal, a humanidade das decisões e a presença do humano que possa de alguma forma avaliar as singularidades de cada caso concreto são fatores fundamentais no Estado Democrático de Direito.

O quadro abaixo traz os resultados coletados em cada etapa.

Quadro 8 – Resultado da triangulação

	LITERATURA	DIRETRIZES	ATORES
Princípios mínimos	“Transparência”, “responsabilidade”, “equidade”, “prestação de contas”, “explicabilidade” e “sustentabilidade” (acrônimo em português TREPES).	Preveem o TREPES, mas não há certificação/órgão de controle.	Ainda não sabem como concretizar.

Preocupações éticas	Proteção de dados e privacidade, vieses discriminatórios e opacidade.	Proteção de dados e privacidade consideradas na LGPD; vieses discriminatórios e opacidade.	Proteção de dados e privacidade e dificuldade em obtenção e estruturação dos dados; vieses discriminatórios como algo humano e opacidade como <i>tradeoff</i> .
Questões sensíveis	“Preditiva”, “julgamento”, “juiz” e “criminal”.	Imprecisão na possibilidade de aplicação de IA na área penal	Impossibilidade de decisões automáticas no Direito.

Fonte: elaboração própria

O que é consenso, afinal, sobre o que é IA ética aplicada ao Direito? Há um padrão de comportamento para os atores envolvidos na criação desses projetos?

O atual conteúdo ético mínimo da IA já existe e é válido para todas as áreas em que a IA é aplicada: são os princípios TREPES. Cada setor específico, entretanto, traz seus desafios e riscos e por isso é necessário ir além da “Era dos princípios”. Mais do que IA uma ética, é fundamental uma IA responsável e a proposição de *design* ético regulado (BODDINGTON, 2017; DIGNUM, 2019).

O consenso sobre o mínimo ético existe. Falta, então, uma regulação geral imposta pelo Poder Público que traga segurança jurídica, atribua responsabilidades, estabeleça órgão(s) de controle e critérios para relatórios de impacto dos projetos, registro de eventos, bem como, promova o letramento cidadão em IA. A cada órgão setorial, por sua vez, cabe o controle e determinação das especificidades e a atribuição dos níveis de risco.

No âmbito do Direito, objeto dessa pesquisa, os riscos apontados no Capítulo 3.1.1 foram: o uso de dados para aprendizado de máquina; privacidade; proteção de dados pessoais, vieses, opacidade/Caixas-pretas/Explicabilidade, transparência, “decisionismo tecnológico” e perda de postos de trabalho e desumanização, que também são preocupações gerais em relação à IA.

No entanto, tais questões podem levar a dificuldades específicas à aplicação da IA ao Direito, como: óbices ao acesso à informação processual; falta de estruturação dos dados judiciais; a polêmica sobre uso de dados sensíveis e onde armazenar dados

coletados; a questão do *open justice*⁸⁰ e a publicidade de atos processuais; o princípio da não discriminação e a igualdade; o princípio da recorribilidade e da ampla defesa e contraditório em face às decisões judiciais que usam algum tipo de IA; a fundamentação das decisões judiciais; a publicidade e controle da administração judicial; o acesso à informação; a eventual formação de precedentes inadequados ao caso concreto; a imparcialidade e independência judiciais; a substituição de tarefas dos advogados e juízes por funcionalidades mais precisas de IA.

Todos esses riscos específicos geram o desafio para o CNJ de criação da “autorregulação regulada”, isso é, de, a partir da legislação geral, estabelecer os níveis de risco da IA aplicada ao Direito, assim como instrumentos de controle, governança e limitação não excessiva dessa tecnologia. Falta uma agenda propositiva de investimentos na integração das ferramentas de processo eletrônico para que se possa viabilizar da melhor forma as aplicações de IA desenvolvidas.

Restou evidente também que o mínimo ético de princípios que atualmente existem, representou grande avanço, mas não é suficiente e não está claro em muitos aspectos práticos. Criar “um código ético de IA” (BODDINGTON, 2017) é tornar os mais de 47 projetos de IA existentes atualmente nos tribunais brasileiros auditáveis e em consonância com os preceitos constitucionais, processuais e que respeite os jurisdicionados. Como?

[...] eu não tenho uma resposta de que a gente vai partir. Até para pensar regulação a gente precisa pensar como que será essa regulação para você não tirar também. Pensando assim parece que quando a gente fala regular só tem só tem dois agentes, dois personagens: o privado e o público. Não! E as pessoas que estão ali, que é a maior quantidade, que estão sofrendo os maiores impactos? Eu não vejo a discussão da regulação chegar nas pessoas da sociedade civil (E3)

Antes que seja tarde e a IA vire um novo Frankenstein, isso é, uma tecnologia que após criada se volte contra o homem (COECKELBERGH, 2020) e automatize desigualdades já existentes, urge esforço coletivo em prol de um desenho ético da IA aplicada ao Direito.

⁸⁰ Que, como se afirmou no Tópico 3.1.2 se traduz na adoção de sistemas interoperáveis de processos eletrônicos, que facilitem a gestão e comunicação dos dados, por entender seu valor na sociedade informacional, inclusive como garantidor da transparência.

CONCLUSÕES

Essa pesquisa buscou responder à seguinte questão: **que parâmetros éticos são identificados e podem ser considerados na literatura, nas diretrizes regulatórias e na percepção de especialistas no desenvolvimento da IA ao Direito?**

Para tanto o percurso da pesquisa foi, em primeiro lugar, descrever o conceito de IA usado como base para o trabalho e as possibilidades de aplicação da IA ao campo do Direito. A IA passou por fases de inverno, nas quais houve pouca produção e evolução, mas atualmente tem passado por crescimento contínuo de desenvolvimento em todas as áreas do conhecimento humano.

Embora o termo seja usado para definir um conjunto de modelos possíveis, o foco dessa pesquisa foi dado à tecnologia de aprendizado de máquina, a qual obteve maiores avanços recentes e que é a mais desenvolvida nas aplicações de IA no Direito (objeto central desse trabalho). Nesse sentido, foram apontados os possíveis usos do aprendizado de máquina no Direito, resumidos no Quadro 1 da tese, a saber: criação de modelos preditivos, organização de documentos, busca textual, automação de documentos e decisões automatizadas. Esses usos, sem dúvida, têm potencial de tornar o sistema de justiça brasileiro mais rápido, mais coerente e com menor custo para a sociedade, isso é, consoante com os ditames trazidos pelo CPC/15. No âmbito profissional, emergem novas profissões jurídicas, em especial ligadas à área de negócios e análise de dados, assim como a tradicional advocacia vê-se obrigada a se adaptar e oxigenar-se.

Ainda que por vezes seja usada como “coringa intelectual”, ou como instrumento sem finalidades muito bem definidas (sob o argumento frágil de inovação pela inovação), é fato que a IA passa por “momento inescapável” de sua aplicação no Direito, em especial o brasileiro. Nesse sentido, foram identificados, no Quadro 3 dessa pesquisa, diversos projetos de IA em andamento no Brasil que demonstraram que o uso da IA tem sido notoriamente estimulado pelo CNJ e tribunais do país e tem se adaptado bem à realidade processual brasileira, que possui um sistema judicial moroso e sobrecarregado de processos. Nesse sentido, é possível concluir que de fato os investimentos em inovação judicial têm aumentado, o que traz a demanda de profissionais adaptados à essa nova realidade.

Sabe-se, entretanto, que esse estímulo não pode vir desacompanhado de limitações responsáveis (sem excessos que se mostrem empecilho ao desenvolvimento dessa tecnologia), eis que as aplicações de IA geram desafios éticos importantes. A partir dessa compreensão, também se defende que o *design* ético da IA deve ter caráter de direito fundamental. Mas quais foram essas balizas e questões éticas apontadas pela literatura, pelas diretrizes regulatórias e pelos especialistas em IA?

Os *inputs* da literatura apresentados no Capítulo 2 apontam que há um consenso mínimo em torno da principal produção científica na área de IA e ética: os princípios norteadores da “transparência”, “responsabilidade”, “equidade”, “prestação de contas”, “explicabilidade” e “sustentabilidade” (acrônimo em português TREPES). A literatura também trouxe as principais preocupações de externalidades negativas que apareceram nas pesquisas. As palavras “viés”, “privacidade”, “dados”, “caixa-preta”, “preditiva”, “armas”, “julgamento”, “juiz” e os tópicos “criminal”, “consumidor”, “civil”, “constitucional”, “governo” e “administração” foram os mais recorrentes, indicando áreas do Direito em que há maior inquietação.

O Capítulo 3 realizou comparação entre as diretrizes regulatórias da Comissão Europeia para Eficiência da Justiça e do Conselho Nacional de Justiça para controle ético de IA aplicada ao Direito, com o objetivo de extrair os indicadores dos sistemas de normatização, confrontando no que se complementam e no que são comuns ou no que divergem, além de apontar desafios que essas normativas enfrentam.

A partir da comparação, foi possível trazer *inputs* mínimos das diretrizes: tal qual a literatura, em comum foi possível encontrar os princípios do respeito aos direitos fundamentais, da não discriminação, da qualidade e segurança, da transparência e do controle do usuário. A área do Direito que mais apresenta risco de aplicação da IA para ambas as diretrizes é o Direito Penal.

Como desafios comuns apareceram: compreender o que é uma aplicação de IA que respeite os direitos fundamentais; quais os parâmetros para interpretabilidade aceitável; como avaliar resultados discriminatórios; a necessidade de contratar equipes plurais para o desenvolvimento dos projetos e seleção criteriosa da base de dados e de criação de certificações que norteiem a avaliação de inteligibilidade dos modelos.

O Capítulo 4 trouxe os *inputs* dos especialistas. Os atores especialistas em IA, por sua vez, mostraram-se inseguros com as normativas existentes atualmente, as quais além de não atribuírem responsabilidades não são claras sobre a aplicação e execução prática dos princípios (por exemplo, não se sabe o que é a explicabilidade e como

concretizá-la). A ética na IA ainda, portanto, tem como ponto comum pretensões principiológicas.

As preocupações éticas dos atores entrevistados foram comuns à literatura, às diretrizes e aos respondentes e giraram em torno da proteção de dados e privacidade, dos vieses discriminatórios e da opacidade algorítmica. Em especial no âmbito da aplicação de IA ao Direito, essas preocupações foram refletidas em três categorias criadas a partir das respostas nas entrevistas: dadaísmo, vieses como limonada pronta e opacidade e *Tradeoffs*.

Os especialistas confirmaram que as diretrizes atualmente existentes são insuficientes, na medida em que compreendem que existe um caos na área de captura e utilização dos dados judiciais. Quanto aos vieses discriminatórios, foi comum o entendimento de que se tratam de questões sociais profundas e muito mais complexas de serem resolvidas que por meio da tecnologia. Nesse sentido, há que se considerar o contexto e realidade brasileira de desigualdades sociais advindas de uma cultura escravocrata, discriminatória e de um passado ditatorial recente (para citar apenas alguns fatores). Ainda no que tange à opacidade algorítmica, os modelos de aprendizagem profunda impõem paradoxos para a sua efetiva regulação e concretização.

E ao final, quais foram os *outputs*? E agora, José?

A hipótese inicial da pesquisa era de que havia um mínimo ético para IA aplicada ao Direito. A partir da triangulação dos resultados, pôde-se compreender que há parâmetros mínimos de controle ético da IA. Entretanto, esse mínimo é comum à IA aplicada em todas as áreas e trata-se dos princípios da TREPES. Não se descarta o avanço que representam os princípios gerais para todas as aplicações em IA, mas apenas se reconhece aqui que eles não são suficientes. Nesse aspecto, o papel do Direito no desenvolvimento de aplicações em IA é criar regulações efetivas, capazes de atribuir responsabilidades sem, no entanto, impedir os avanços dessa tecnologia que, bem aplicada, tem o potencial de melhorar o acesso a uma justiça eficiente.

A IA não é magia. Há muito que o ser humano possa fazer para controlar e limitar suas externalidades negativas. Aqui, a partir de toda a pesquisa, recomenda-se como principal *output* o modelo de “autorregulação regulada”. Isso é, um modelo em que cabe ao Estado impor padrões e parâmetros mínimos vinculantes de governança gerais e aos setores em que a IA é aplicada e criar diretrizes que tragam uma abordagem baseada em riscos. Afinal, como se demonstrou, cada área possui seus riscos e pontos críticos sobre os quais deve se debruçar. E embora o Brasil possua atualmente a EBIA e

um Projeto de Lei geral que demonstram interesse do país em trazer balizas na área, essas diretrizes ainda se mostram insuficientes na prática.

No Direito, por exemplo, foram pontuados como maiores riscos de uso da IA a área penal e em decisões automatizadas sem revisão humana. Nesse aspecto, embora a IA possa trazer uniformidade decisória e eficiência jurisdicional é fundamental que as diretrizes setoriais repensem a necessidade recursal, possibilitando-se a revisão das decisões automatizadas; regulem a atuação de empresas desinteressadas na causa em ODRs que utilizam IA; priorizem a paridade de armas processuais no uso da jurimetria; a transparência e equidade processuais nos sistemas de IA aplicadas ao processo e o respeito aos direitos humanos.

A regulação e as diretrizes setoriais precisam, antes de tudo, dialogar com a tecnologia e suas limitações. Além disso, recomenda-se que identifiquem que para alguns problemas trazidos pela aplicação de IA, a correção vem por meio de maior esforço tecnológico e para outros, maior esforço coletivo, capaz de aprofundar a construção de medidas que tornem factíveis os princípios éticos apontados por essa pesquisa. Assim, uma possível sugestão é a de que todos os projetos precisam de ampla documentação dos eventos, relatórios de análise de impactos e gestão de dados que evite sobre/sub-representação, auditorias contínuas e que se realizem constantes testes de segurança. Também se sugere a criação de conselhos de ética por projeto, que assegurem compromisso com a governança e controlem de forma externa as aplicações mais arriscadas de IA. São fundamentais também investimentos na área de educação tanto dos novos profissionais do mercado jurídico, quanto dos desenvolvedores dos projetos de IA, para que se crie a consciência cidadã acerca das externalidades negativas de cada modelo.

A conclusão também é o momento de reconhecimento de uma das principais limitações da pesquisa: a compreensão de que existem questões éticas profundas arraigadas na sociedade, cujas regras não dão conta de controlar, como é o caso do viés discriminatório (que pode ser reforçado pela IA).

Por outro lado, não há dúvidas que para mudança cultural e estrutural profunda é fundamental o prosseguimento das investigações no tema, que busquem propiciar o desenvolvimento ético da IA, humanocêntrica. E que, assim, propicie um ambiente seguro para investimentos na área, assim como atenda às necessidades dos cidadãos e jurisdicionados brasileiros.

REFERÊNCIAS

- ADEODATO, J. M. **Ética e Retórica: para uma teoria da dogmática jurídica**. São Paulo: Saraiva, 2002.
- ALARIE, B.; NIBLETT, A.; YOON, A. H. How artificial intelligence will affect the practice of law. **University of Toronto Law Journal**, v. 68, p. 106–124, 2018.
- ALTENHOFER, M. et al. (EDS.). The Valuable Citizens of Smart Cities: The Case of Songdo City. In: **Changing Worlds Ideologies, Utopias and Ambitions in Science & Technology**. [s.l.] Graduate Journal of Social Science, 2016. v. 12.
- ARBIX, G. A transparência no centro da construção de uma IA ética. **Novos Estudos CEBRAP**, v. 39, n. 2, p. 395–413, 2020.
- ASIMOV, Isaac. **Eu, Robô** (1950). Trad. Aline Storto Pereira. São Paulo: Aleph, 2014.
- ATIENZA, M. **Filosofía del derecho y transformación social**. Madri: Trotta, 2017.
- BAPTISTA, P.; KELLER, C. I. Por que , quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. n. 2011, p. 123–163, 2016.
- BARDIN, L. **Análise de Conteúdo**. Lisboa: Edições 70, 1977.
- BAROCAS, S.; SELBST, A. D. Big Data ' s Disparate Impact. v. 671, p. 671–732, 2016.
- BARREDO ARRIETA, A. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. **Information Fusion**, v. 58, n. December 2019, p. 82–115, 2020.
- BAUMAN, Zygmunt. **Vida em fragmentos: Sobre a ética pós-moderna**. São Paulo: Zahar, 2011.
- BECKER, Daniel. O acesso à informação jurídica on-line como medida de garantia ao direito de acesso à justiça. In: **Advocacia 4.0** [livro eletrônico] /coordenação Viviane Nóbrega Maldonado e Bruno Feigelson. São Paulo: Thomson Reuters Brasil, 2019.
- BENJAMIN, Ruha. Retomando nosso fôlego: estudos de ciência e tecnologia, teoria racial crítica e a imaginação carcerária. In: **Comunidades, Algoritmos e Ativismos Digitais: olhares afrodiáspóricos**. Org. Tarcízio Silva. 2ª Ed. 1ª Reimp. São Paulo: Literarua, 2021, p. 15-26.
- BIONI, B. R.; LUCIANO, M. O princípio da precaução na regulação de Inteligência Artificial: seriam as leis de proteção de dados o seu portal de entrada? In: **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. In: Ana Frazão; Caitlin Mulholland (Coord.). **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019. p. 207–231.
- BODDINGTON, P. **Towards a Code of Ethics for Artificial Intelligence**. Oxford: Springer, 2017.
- BODÓ, B.; HELBERGER, N.; VREESE, C. H. DE. Political micro-targeting : a Manchurian candidate or just a dark horse ? **Internet Policy Review**, v. 6, n. 4, p. 1–13, 2017.
- BOGDAN, Robert; BIKLEN, Sari. **Investigação qualitativa em educação: uma introdução à teoria e aos métodos**. Porto: Porto Editora, 1994.

- BONI, V.; QUARESMA, J. Aprendendo a entrevistar como fazer entrevistas em Ciências Sociais. **Em Tese**, v. 2, n. 1, p. 68–80, 2005.
- BORENSTEIN, J.; HERKERT, J.; MILLER, K. Self-Driving Cars: Ethical Responsibilities of Design Engineers. n. june, p. 67–75, 2017.
- BRINK, H.; RICHARDS, J. W.; FETHEROLF, M. Real-World Machine Learning. Meap. Manning, 2016.
- BUCKLEY, B.; HUNTER, M. Say cheese ! Privacy and facial recognition. **Computer Law & Security Review**, v. 27, n. 6, p. 637–640, 2011.
- CASTELLS, Manuel. **A Era da informação: economia, sociedade e cultura**. Vol I. A Sociedade em Rede. 14 reimp. Trad. Roneide Venancio Majer. São Paulo: Paz e Terra, 2011.
- CASTRO, M. A. **Personalidade jurídica do robô e sua efetividade no Direito**. 222f. 2009. Tese (Doutorado em Direito Público). Universidade Federal da Bahia, Salvador, 2009.
- CASSINO, João Francisco; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu. **Colonialismo de Dados: Como Opera a Trincheira Algorítmica na Guerra Neoliberal**. São Paulo: Autonomia Literária, 2021.
- CERVANTES, J. A. et al. **Artificial Moral Agents: A Survey of the Current Status**. [s.l.] Springer Netherlands, 2019.
- CESSNA, A. Scholarship Cornell Law: A Digital Repository Future of AI and Law. 2015.
- CHEN, H. et al. Next Generation Technology for Epidemic Prevention and Control : Data-Driven Contact Tracking. **IEEE Access**, v. 7, n. January 2015, p. 2633–2642, 2019.
- COECKELBERGH, M. Responsibility and the Moral Phenomenology of Using Self-Driving Cars Responsibility and the Moral Phenomenology of Using Self-Driving Cars. **Applied Artificial Intelligence**, v. 30, n. 8, p. 748–757, 2016.
- COLLINGRIDGE, D. **The Social Control of Technology**. Londres: Frances Pinter, 1980.
- CRANE, D. A. Search Neutrality as an Antitrust Principle. v. 5, n. 5, 2012.
- CRESWELL, John W. **Projetos de pesquisa: Métodos Qualitativo, Quantitativo e Misto**. Trad. Magda França Lopes. 3. ed. São Paulo: Artmed, 2010.
- CUEVA, Ricardo Villas Bôas. Inteligência Artificial no Judiciário. In: Inteligência Artificial e Direito Processual: **Os Impactos da Virada Tecnológica no Direito Processual**. Coord. Dierle Nunes, Paulo Henrique dos Santos Lucon, Erick Navarro Wolkart. Salvador: Juspodvim, 2021, p. 79-91.
- CUNHA, Leonardo Carneiro Da. **A Fazenda Pública em Juízo**. 18ª Ed., Rio De Janeiro: FORENSE, 2021.
- DALY, A. et al. **Artificial Intelligence Governance and Ethics : Global Perspectives**. n. June, p. 20–21, 2019.
- DIGNUM, V. **Responsible Artificial Intelligence: HowtoDevelop and Use AI in a Responsible Way**. Cham: Springer Netherlands, 2019.

- DOMINGOS, P. A Few Useful Things to Know About Machine Learning. *Communications of the ACM*, v. 55, p. 78–87, 2012.
- DOMINGOS, P. **O Algoritmo Mestre: Como a Busca Pelo Algoritmo de Machine Learning Definitivo Recriará Nosso Mundo**. São Paulo: Novatec, 2017.
- DWORK, C.; ROTH, A. The Algorithmic Foundations of Differential Privacy. v. 9, p. 211–407, 2014.
- EBERS, M. Regulating AI and Robotics: Ethical and Legal Challenges. In: EBERS, M.; NAVARRO, S. N. (Eds.). **Algorithms and Law**. [s.l.] Cambridge University Press, 2019. v. 2019p. 1–51.
- EN, B.; PÖLL, M. Are you (self-)tracking? Risks, norms and optimisation in self-quantifying practices. *Graduate Journal of Social Science*, v. 12, p. 37–57, 1 maio 2016.
- FALEIROS JR., José Luiz de Moura. Breves reflexões sobre os impactos jurídicos do algoritmo GPT-3. In: Direito digital e inteligência artificial: diálogos entre Brasil e Europa. Coord. Felipe Braga Neto (et al.) Indaiatuba: Editora Foco, 2021, p. 521-531.
- FARRANHA, Ana Claudia; BATAGLIA, Murilo. Algoritmos, Administração Pública e Judiciário: Reflexões Advindas do Projeto Constituição Anotada. In: **Direito e Tecnologia: um debate multidisciplinar**. Org. Murilo Siqueira Comério e Tainá Aguiar Junquilha. Rio de Janeiro: Lumen Juris, 2020, p. 199-220.
- FERRARI, Isabela; BECKER, Daniel. Direito à explicação e decisões automatizadas: reflexões sobre o princípio do contraditório. In: Inteligência Artificial e Direito Processual: **Os Impactos da Virada Tecnológica no Direito Processual**. Coord. Dierle Nunes, Paulo Henrique dos Santos Lucon, Erick Navarro Wolkart. Salvador: Juspodvim, 2021, p. 277-303.
- FLORIDI, Luciano. **The 4th Revolution: how infosphere is reshaping human reality**. Oxford: Oxford University Press, 2014.
- FLORIDI, L. et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, v. 28, n. 4, p. 689–707, 2018.
- FLORIDI, L. What the Near Future of Artificial Intelligence Could Be. p. 1–15, 2019.
- FLORIDI, L. AI and Its New Winter : from Myths to Realities. p. 1–3, 2020.
- FLORIDI, L. et al. How to Design AI for Social Good : Seven Essential Factors. *Science and Engineering Ethics*, v. 26, n. 3, p. 1771–1796, 2020.
- FLORIDI, L.; COWLS, J. A United Framework of Five Principles for AI in Society. p. 1–15, 2019.
- FRAZÃO, Ana. Responsabilidade civil de administradores de sociedades empresárias por decisões tomadas com base em sistemas de inteligência artificial. In: Ana Frazão; Caitlin Mulholland (Coord.) **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019, p.481-521.
- FRAZÃO, Ana. **Decisões algorítmicas e direito à explicação** (2021a). Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/decisoes-algoritmicas-e-direito-a-explicacao-24112021>. Acesso em: 15 dez. 2021.
- GARWOOD, M. Artificial Intelligence. *Twice*, v. 6, p. 10–11, 2018.

- GOMES, R. D. DE P. Carros Autônomos e os desafios impostos pelo ordenamento jurídico: uma breve análise sobre a responsabilidade civil envolvendo veículos inteligentes. In: *Inteligência Artificial e Direito: ética, regulação e responsabilidade*. In: Ana Frazão; Caitlin Mulholland (Coord.). **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019. p. 567–585.
- GIUFFRIDA, Iria; LEDERER, Fredric; VERMEYS, Nicolas. **A legal perspective on the trials and tribulations of AI**: how artificial intelligence, the internet of things, smart contracts, and other technologies will affect the law. *Case Western Reserve Law Review*, v. 68, n. 3, p. 747–782, 2018.
- GONÇALVES, L. R. **A tutela jurídica de trabalhos criativos feitos por aplicações de inteligência artificial no Brasil**. [s.l.] Universidade Federal do Paraná, 2019.
- GORDON-MURNANE, L. Ethical, Explainable Artificial Intelligence: Bias and Principles. **Online Searcher**, v. 42, n. 2, p. 22–30, 2018.
- GRAVES, R. **Os mitos gregos. Trad. Fernando Klabin v. 1 e 2**. Rio de Janeiro: Nova Fronteira, 2018.
- HAN, B.-C. **No exname: Perspectivas do digital**. Petrópolis: Vozes, 2019.
- HAO, X.; ZHANG, G. Technical Survey Deep Learning. **International Journal of Semantic Computing**, v. 10, n. 3, p. 417–439, 2016.
- HEAWOOD, J. Pseudo-public political speech: **Democratic implications of the Cambridge Analytica scandal**. v. 23, n. June, p. 429–434, 2018.
- HOVEN, Jeroen van den; VERMAAS, Pieter E.; POEL, Ibo van de. **Handbook of Ethics, Values and Technological design**: sources, theory, values and application domains. Nova Iorque: Springer, 2015.
- HOYBJERG, B. P.; BUCK, A. Established Industries and Legal Practices Self-Driving Automobiles. v. 36, n. 5, 2019.
- ITS. Planos estratégicos de desenvolvimento de Inteligência Artificial. 2020.
- JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 2015.
- JORDAN, Michael. **Artificial Intelligence — The Revolution Hasn't Happened Yet**. 2018. Disponível em: <https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7>. Acesso em: 08 nov. 2021.
- JUNQUILHO, T. A.; JEVEAUX, G. C. “TUPI, OR NOT TUPI”: A NECESSÁRIA E DEFINITIVA ADAPTAÇÃO DA TEORIA DOS PRECEDENTES AO BRASIL. **Revista da Faculdade Mineira de Direito**, v. 20, n. 39, p. 130–145, 28 dez. 2017.
- KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R., **Noise: A Flaw in Human Judgment**. Little, Brown Spark: Nova Iorque, 2021.
- KATZ, D. M. QUANTITATIVE LEGAL PREDICTION — OR — HOW I LEARNED TO STOP WORRYING AND START PREPARING FOR THE DATA-DRIVEN FUTURE OF THE LEGAL SERVICES INDUSTRY Welcome to law ’ s information revolution 1 — revolution already in progress . 2 While the 2008 financial cri. **Emory Law Journal**, v. 62, n. July 2011, p. 909–966, 2012.
- KAUFMAN, Dora. Inteligência Artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. **PAULUS: Revista de**

Comunicação da FAPCOM. São Paulo, v. 5, n. 9, jan./jul. 2021.

KAUFMAN, D. Os meandros da Inteligência Artificial : conceitos-chave para leigos. **Estadão**, p. 1–10, 2018.

KAY, J. AI and Education : Grand Challenges. p. 12–15, 2012.

KEARNS, M.; ROTH, A. **The Ethical Algorithm: The Science of Socially Aware Algorithm Design.** Oxford: Oxford University Press, 2020.

KIETZMANN, J. et al. ScienceDirect Deepfakes : Trick or treat ? **Business Horizons**, v. 63, n. 2, p. 135–146, 2020.

KOIVISTO, I. **The anatomy of transparency: the concept and its multifarious implications.** EUI working papers, Fiesole, set. 2016. Disponível em: <http://hdl.handle.net/1814/41166>. Acesso em: 6 abr. 2021.

KRIPPENDORFF, Klaus. **Content Analysis: An Introduction to Its Methodology.** 4ed. Londres: Sage Publications, 2018.

LARSSON, S.; HEINTZ, F. Transparency in artificial intelligence. v. 9, n. 2, p. 1–16, 2020.

LAUTERBACH, A. Artificial intelligence and policy : quo vadis ? **Digital policy, regulation and governance**, v. 21, n. 3, p. 238–263, 2019.

LAWTON, G. Using AI to Make Games Smarter. **IEEE Computer Society**, p. 5–9, 2011.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436–444, 2015.

LEE, K.-F. **Inteligência Artificial: como os robôs estão mudando o mundo, a forma como amamos, nos comunicamos e vivemos.** Trad. Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.

LIMA, Herisandro da Silva. **As práticas da liderança em uma empresa do setor de varejo da Paraíba.** 2017. 133 f. Dissertação (Mestrado em Administração) – Programa de Pós-Graduação em Administração, Universidade Federal da Paraíba, João Pessoa, 2017.

LIN, Patrick. **The ethics of saving lives with autonomous cars are far murkier than you think.** Disponível em: <<http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars>>. Acesso em: 30 jun. 2020.

LINS E HORTA, R. de. **Direito e tomada de decisão:** Elementos para uma Teoria da Decisão Jurídica. Universidade de Brasília, 2019, 177p.

LINS E HORTA, R. de; COSTA, A. A. Vieses na decisão judicial e desenho institucional: uma discussão necessária na era da pós-verdade. **Cadernos Adenauer**, v. XVIII, n. 1, p. 11–34, 2017.

LOH, Stanley. **Volume, Velocidade, Variedade, Veracidade e Valor:** Como os 5 Vs do Big Data estão impactando as Organizações e a Sociedade. Porto Alegre: Edição do Autor, 2019.

LOURIDAS, P.; EBERT, C. Machine Learning. **IEEE Software**, p. 110–115, 2016.

MAGRANI, E. **A internet das coisas.** Rio de Janeiro: FGV Editora, 2018.

MAGRANI, E. **Entre dados e robôs: ética e privacidade na era da**

hiperconectividade. Porto Alegre: Arquipélago Editorial, 2019.

MAIA FILHO, M. S.; JUNQUILHO, T. A.. INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO: LIÇÕES DO PROJETO VICTOR. **Revista Humanidades e Inovação**, v. 8, n. 48, p. 147–160, 2021.

MANHEIM, K.; KAPLAN, L. **Artificial Intelligence**: Risks to Privacy and Democracy. v. 106, p. 106–188, 2019.

MANZINI, E. J. Entrevista semi estruturada estudo. UNESP, Marília., 2018.

MARCHANT, Gary E.. **Artificial Intelligence and the Future of the Legal Practice**. Disponível em:

<<https://www.americanbar.org/content/dam/aba/administrative/litigation/materials/2017-2018/2018-sac/written-materials/artificial-intelligence-and-the-future.authcheckdam.pdf>>. Acesso em: 30 mai. 2021.

MCCARTHY, J. et al. A proposal for the Dartmouth summer research project on artificial intelligence. **AI Magazine**, v. 27, n. 4, p. 12–14, 2006.

MERCIER, H.; SPERBER, D. Why do humans reason? Arguments for an argumentative theory. **Behavioral and Brain Sciences**, v. 34, p. 57–111, 2011.

MESQUITA, C. T. Inteligência Artificial e Machine Learning em Cardiologia – Uma Mudança de Paradigma. **International Journal of Cardiovascular Sciences**, v. 30, n. 3, p. 187–188, 2017.

MIRANDA, E. R.; WILLIAMS, D. Artificial intelligence in organised sound. **Organised Sound**, v. 20, n. 1, p. 76–81, 2015.

MORAV, M. et al. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. **Science**, v. 356, n. May, p. 508–513, 2017.

MORLEY, J. et al. Ethical guidelines for COVID-19 tracing apps. **Nature Machine Intelligence**, v. 582, p. 29–31, 2020.

MORROW, D. R. When Technologies Makes Good People Do Bad Things : Another Argument Against the Value- Neutrality of Technologies. p. 329–343, 2014.

MOU, Y.; XU, K. The media inequality: Comparing the initial human-human and human-AI social interactions. **Computers in Human Behavior**, v. 72, n. 2017, p. 432–440, 2017.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z.. Inteligência Artificial e a Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o Direito à explicação perante a tomada de decisões por meio de machine learning. In: Ana Frazão e Caitlin Mulholland(Coord.). **Inteligência Artificial e Direito: ética, regulação e responsabilidade**. 2019, p. 265-290.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z.. Entre as leis da robótica e a ética: regulação para o desenvolvimento da inteligência artificial. In: **Direito digital e inteligência artificial**: diálogos entre Brasil e Europa. Coord. Felipe Braga Neto (et al.) Indaiatuba: Editora Foco, 2021, p. 65-80.

NECHUSHTAI, E.; LEWIS, S. C. Computers in Human Behavior What kind of news gatekeepers do we want machines to be ? Filter bubbles , fragmentation , and the normative dimensions of algorithmic recommendations. **Computers in Human Behavior**, v. 90, n. June 2018, p. 298–307, 2019.

NEFF, G.; NAGY, P. Talking to Bots : Symbiotic Agency and the Case of Tay. v. 10, p.

4915–4931, 2016.

NUNES, Dierle; MALONE, Hugo. O uso da tecnologia na prevenção efetiva de conflitos. In: *Inteligência Artificial e Direito Processual: Os Impactos da Virada Tecnológica no Direito Processual*. Coord. Dierle Nunes, Paulo Henrique dos Santos Lucon, Erick Navarro Wolkart. Salvador: Juspodvim, 2021, p. 123-145.

O'NEIL, Cathy. **Weapons of math destruction: how big data increases inequality and threatens democracy**. Nova York: Crown Publishers, 2016.

PARISER, E. **The Filter Bubble: What the Internet Is Hiding from You**. Nova Iorque: The Penguin Press, 2015.

PARK, S. H.; HAN, K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis. v. 286, n. 3, 2018.

PASQUALE, F. **The black box society: the secret algorithms that control money and information**. Cambridge: Harvard University Press, 2015.

PEIXOTO, Fabiano Hartmann. **Inteligência Artificial e Direito: Convergência Ética e Estratégica**. Curitiba: Alteridade, 2020.

PEREZ, J. A. et al. Artificial Intelligence and Robotics. e **UK-RAS White Paper Series on Robotics and Autonomous Systems (RAS)**., 2017.

QIAN, W. et al. Simulating Chalk Art Style Painting. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 31, n. 12, p. 1–20, 2017.

RASCHKA, S. Single-Layer Neural Networks and Gradient Descent. 2015. Disponível em: <https://sebastianraschka.com/Articles/2015_singlelayer_neurons.html>. Acesso em: 01 nov. 2020

RIBEIRO, M. et al. Visão baseada em recursos: uma análise bibliométrica dos últimos 11 anos. **Revista de Ciências da Administração**, v. 14, n. 34, p. 39–59, 2012.

ROSA, Alexandre Moraes da; GUASQUE, Bárbara. O avanço da disrupção nos tribunais brasileiros. In: *Inteligência Artificial e Direito Processual: Os Impactos da Virada Tecnológica no Direito Processual*. Coord. Dierle Nunes, Paulo Henrique dos Santos Lucon, Erick Navarro Wolkart. Salvador: Juspodvim, 2021, p. 93-121.

RULE, Colin. **Online Dispute Resolution for Business**. B2B, E-Commerce, Consumer, Employment, Insurance, and Other Commercial Conflicts. São Francisco: Jossey-Bass, 2002

RUSSELL, S. **Human Compatible: Artificial Intelligence and the problem of control**. New york: Viking, 2019.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: a modern approach**. New Jersey: Alan Apt, 2013.

SADATRASOUL, S. M. et al. Credit scoring in banks and financial institutions via data mining techniques : A literature review. v. 1, n. 2, p. 119–129, 2013.

SAMATAS, M. “Austerity Surveillance ” in Greece under the Austerity Regime. **Gogitatio**, v. 3, n. 3, p. 68–80, 2015.

SAMPAIO, Rafael; LYCARIÃO, Diógenes. **Análise de conteúdo categorial: manual de aplicação**. Disponível em: https://repositorio.enap.gov.br/bitstream/1/6542/1/Analise_de_conteudo_categorial_fina

[L.pdf](#). Acesso em: 15 dez. 2021.

SANDEL, M. **Justiça: o que é fazer a coisa certa?** Rio de Janeiro: Civilização Brasileira, 2009.

SCHNEIDER, S. **Fiction and philosophy: from the travel to superintelligence.** Oxford: Blackwell Publishing Ltd, 2009.

SHARMA, G. D.; YADAV, A.; CHOPRA, R. Artificial intelligence and effective governance: A review, critique and research agenda. **Sustainable Futures**, v. 2, Nov., 2020.

SILVA, N. C. DA. Inteligência Artificial. In: (Coord.) Ana Frazão e Caitlin Mulholland. **Inteligência Artificial e Direito: ética, regulação e responsabilidade.** São Paulo: Revista dos Tribunais, 2019. p. 35–52.

SILVEIRA, Sergio Amadeu de. **Tudo sobre tod@s:** Redes digitais, privacidade e venda de dados pessoais. São Paulo: Edições SESC, 2017.

SLEE, T. **Uberização: a nova onda do trabalho precarizado.** São Paulo: Elefante, 2017.

SOARES, M. N.; KAUFFMAN, M. E.; CHAO, K.-M. Inteligência Artificial : Impactos no Direito e na Advocacia. **RDP**, v. 17, n. 93, p. 104–133, 2020.

SOUZA, C. A. P.; OLIVEIRA, J. V. Sobre os ombros de robôs? A Inteligência Artificial entre fascínios e desilusões. In: Ana Frazão e Caitlin Mulholland(Coord.). **Inteligência Artificial e Direito: ética, regulação e responsabilidade.** São Paulo: Revista dos Tribunais, 2019. p. 65–81.

SURDEN, H. Machine learning and law. **Washington Law Review**, v. 89, n. 1, p. 87–115, 2014.

SUSSKIND, Richard. **Online Courts and the future of Justice.** Oxford: Oxford University Press, 2019.

SUSSKIND, Richard. **Tomorrow’s Lawyers: An Introduction to Your Future.** Oxford: Oxford University Press, 2017.

TAYLOR, R. D. Quantum Artificial Intelligence: A “precautionary” U.S. approach? **Telecommunications Policy**, August 2019, p. 1-12, 2020.

TECUCI, G. Artificial intelligence. **Wiley Periodicals**, v. 4, n. April, p. 168–180, 2012.

TROMPOUKIS, C.; KOUKOUTAS, D. Greek mythology : the eye , ophthalmology , eye disease, and blindness. **Canadian Journal of Ophthalmology**, v. 42, n. 3, p. 455–459, 2007.

TZARA, Tristan. **Manifesto Dadá.** (1918). Disponível em: <<http://docplayer.com.br/39091112-Manifesto-dadaista-de-tristan-tzara-de-1918.html>>. Acesso em: 18 nov. 2021.

TURING, A. M. Computing machinery and intelligence. **Mind**, v. LIX, p. 433–460, 1950.

TZIRULNIK, V.; BOAVENTURA, E. Uma indústria em transformação: o seguro e a inteligência artificial. In: (Coord.) Ana Frazão e Caitlin Mulholland. **Inteligência Artificial e Direito: ética, regulação e responsabilidade.** São Paulo: Revista dos Tribunais, 2019. p. 523–538.

- VAISHYA, R. et al. Diabetes & Metabolic Syndrome : Clinical Research & Reviews Artificial Intelligence (AI) applications for COVID-19 pandemic. v. 14, p. 337–339, 2020.
- VALENTE; Ana Cláudia de Souza; SUGAI; Kin Modesto; JUNQUILHO; Tainá Aguiar. A Inteligência Artificial no TJDFT: estratégias de publicidade e princípio da transparência. Compolítica. 2021
- VELLINGA, N. E. From the testing to the deployment of self-driving cars : Legal challenges to policymakers on the road ahead. **Computer Law & Security Review: The International Journal of Technology Law and Practice**, v. 33, n. 6, p. 847–863, 2017.
- VERGARA, S. C. Análise de conteúdo como técnica de análise de dados qualitativos no campo da administração: potencial e desafios. **Revista de Administração Contemporânea**, v. 15, n. 4, p. 761–765, 2011.
- VERONESE, A.; SILVEIRA, A.; LEMOS, A. N. L. E. Inteligência Artificial, Mercado Único Digital e a postulação de um direito às inferências justas e razoáveis: uma questão jurídica entre a ética e a técnica. In: **Ana Frazão e Caitlin Mulholland (Coord.). Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Revista do Tribunais, 2019. p. 233–264.
- VODÁ, A. I.; RADU, L.-D. V. Artificial Intelligence and the Future of Smart Cities. **BRAIN – Broad Research in Artificial Intelligence and Neuroscience**, v. 9, n. 2, p. 110–127, 2015.
- WACHTER, S.; MITTELSTADT, B.; FLORIDI, L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. v. 7, n. 2, p. 76–99, 2018.
- WALKER, Joshua. **Edição Brasileira On Legal AI: um rápido tratado sobre a Inteligência Artificial no Direito**. Coord. e Prefácio Alexandre Zavaglia Coelho, Marina Feferbaum, Alexandre Pachecho da Silva. São Paulo: Revista dos Tribunais, 2021.
- WANG, F. F. **Online Dispute Resolution: Technology, Management and Legal Practice from an International Perspective**. Oxford: Chandos Publishing, 2009.
- WATSON, D.; FLORIDI, L. **The Explanation Game: A Formal Framework for Interpretable Machine Learning**. [s.l.] Springer Netherlands, 2020.
- XAVIER, L. P.; SPALER, M. G. Patrimônio de afetação: uma possível solução para os danos causados por sistemas de inteligência artificial. In: **(Coord.) Ana Frazão e Caitlin Mulholland Inteligência Artificial e Direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019. p. 541–562.
- ZHANG, J. et al. Rapid and accurate intraoperative pathological diagnosis by artificial intelligence with deep learning technology. **Medical Hypotheses**, v. 107, n. 37, p. 98–99, 2017.
- ZICKLER, C. et al. Brain painting: Usability testing according to the user-centered design in end users with severe motor paralysis. **Artificial Intelligence in Medicine**, v. 59, n. 2, p. 99–110, 2013.
- ZUBOFF, S. Big other: Surveillance capitalism and the prospects of an information civilization. **Journal of Information Technology**, v. 30, n. 1, p. 75–89, 2015.
- ZUBOFF, S. **The Age of Surveillance Capitalism. The Fight for a Human Future at**

the New Frontier of Power. New york: Public Affairs, 2019.

ZWITTER, A.; GSTREIN, O. J. Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. **Journal of International Humanitarian Action**, v. 5, n. 1, p. 4, 2020.

APÊNDICES

Termo de Consentimento Livre e Esclarecido (TCLE)

Você está sendo convidado a participar da pesquisa “Inteligência Artificial (IA) aplicada ao Direito e a problemática do controle ético: limites, possibilidades e percepções”, de responsabilidade de Tainá Aguiar Junquilha, estudante de doutorado da Universidade de Brasília. O objetivo desta pesquisa é avaliar como os atores envolvidos em projetos de aplicação de IA ao Direito consideram aspectos de controle ético. Assim, gostaria de consultá-lo/a sobre seu interesse e disponibilidade de cooperar com a pesquisa.

Você receberá todos os esclarecimentos necessários antes, durante e após a finalização da pesquisa, e lhe asseguro que o seu nome não será divulgado, sendo mantido o mais rigoroso sigilo mediante a omissão total de informações que permitam identificá-lo/a. Os dados provenientes de sua participação na pesquisa, tais como questionários, entrevistas, fitas de gravação ou filmagem, ficarão sob a guarda do/da pesquisador/a responsável pela pesquisa.

A coleta de dados será realizada por meio de resposta ao questionário realizado, em entrevista gravada, num período aproximadamente de 1 (uma) hora, que se darão no local de preferência do(a) senhor(a). É para estes procedimentos que você está sendo convidado a participar. Sua participação na pesquisa não implica em nenhum risco.

Espera-se com esta pesquisa compreender como a legislação e os atores envolvidos em projetos de aplicação de IA ao Direito consideram aspectos de controle ético.

Sua participação é voluntária e livre de qualquer remuneração ou benefício. Você é livre para recusar-se a participar, retirar seu consentimento ou interromper sua participação a qualquer momento. A recusa em participar não irá acarretar qualquer penalidade ou perda de benefícios.

Se você tiver qualquer dúvida em relação à pesquisa, você pode me contatar através do telefone 61 981851488 ou pelo e-mail taina.aguiarj@gmail.com.

A equipe de pesquisa garante que os resultados do estudo serão divulgados na Universidade de Brasília, podendo ser publicados posteriormente na comunidade científica.

Este projeto foi revisado e aprovado pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da Universidade de Brasília. As informações com relação à assinatura do TCLE ou aos direitos do participante da pesquisa podem ser obtidas por meio do e-mail do CEP/CHS: cep_chs@unb.br ou pelo telefone: (61) 3107 1592.

Este documento foi elaborado em duas vias, uma ficará com o/a pesquisador/a responsável pela pesquisa e a outra com você.

Assinatura do/da participante

Assinatura do/da pesquisador/a

Brasília, ___ de _____ de _____

ANEXO I

Na primeira busca, foram selecionados os descritores “*ethics*” e “*artificial intelligence*”, com o intuito de encontrar aspectos presentes na literatura sobre ética para IA em geral. Foram encontrados 661 artigos e baixados 503 artigos cujo acesso estava disponível. Na Figura 6 observa-se a nuvem das 60 palavras que mais aparecem após o processamento dos textos completos localizados e carregados no *software Orange*, excluindo-se as expressões presentes no Anexo I:

de

e

1

2018

1

en

2

r

2019

https

el

2017

c

al

also

di

may

3

p

n

4

2016

j

10

one
org
www
del
un
f
5
los
b
pp
even
however
con
doi
f
2015
6
u
8
por
v
w
las
two
h
12
2012
il
http
11

ANEXO II

Na segunda busca procurou-se selecionar descritores que envolvessem ética, IA e Direito, para obter informações presentes na literatura recente sobre a discussão da ética em IA aplicada ao campo do Direito. Nesse sentido, elegeram-se os descritores “artificial intelligence” e “*decision making*” e “*Law*”. Foram encontrados 121 resultados e baixados 73 artigos com acesso disponível. Os textos que resultaram da busca foram inseridos no *software Orange* que usa IA para criar métricas. Na figura 10 observa-se a nuvem de palavras que mais aparecem após o processamento dos 73 textos encontrados, excluindo-se as expressões presentes no Anexo II:

t
la
31
28
em
natl
sch
ieee
nov
jan
31
oct
ct
co
09
20
22
j
10th
mar
01
eciair
25
27

kasetsart

2017

2018

2019

17

b

whithout

ANEXO III

Por fim, a terceira busca utilizou os termos “*artificial intelligence*” e “*principle*” e “*Law*”. Encontrou-se 85 resultados e foram baixados 61 arquivos disponíveis. Ao importar os documentos para o *software Orange*, buscou-se averiguar nos textos completos quais as palavras mais recorrentes, que aparecem na Figura 14 (excluindo-se as expressões presentes no Anexo III):

1
issues
para
better
second
art
third
better
25
2012
26
2016
see
note
supra
making
would
com
et
g
new
could
0
4

u
also
j
two
w
therefore
pp
c
o
con
28
ways
example
review
7
research
cid
used
8
9
make
rev
2020
https
whether
la
well
different
time
must
impact
using
states
many

cc
need
id
14
2013
x
vol
approach
29
way
journal
university
whithout
19
13
20
given
set
2014
15
16
tools
rather
often
made
18
22
including
cannot
ing
21
que
k
tion

means

23

17

c

number

non

important

31

relevant

e

2

1

may

n

r

2018

use

de

2017

3

v

one

www

10

system

p

based

5

2019

b

org

6

en

f

2015

h

12

11

pdf

doi

within

el

c

30

o

us

al

perma

might

even

1