REVISTA DIREITO MACKENZIE

ISSN: 23172622

TRANSPARÊNCIA *VERSUS* EXPLICAÇÃO: O PAPEL DA AMBIGUIDADE NA IA JURÍDICA*

Elena Esposito**

RECEBIDO EM:	12.9.2022
APROVADO EM:	17.11.2022

* Artigo originalmente publicado em *Journal of Cross-Disciplinary Research in Computational Law*, v. 1, n. 1, 2021. Tradução de Isabela Gomes Ribeiro (Mestranda em Direito Político e Econômico na Universidade Presbiteriana Mackenzie) e revisão técnica de Izabela Zonato Villas Boas (Mestra em Direito Político e Econômico na Universidade Presbiteriana Mackenzie, Mestra pelo Instituto Internacional de Sociologia Jurídica em Oñati (Espanha), membro do Research Committee on Sociology of Law da International Sociological Association).

** Professora de sociologia na Universidade de Bielefeld, Alemanha, e na Universidade de Modena-Reggio Emilia, Itália. *E-mail*: elena.esposito@uni-bielefeld.de

2022 | v. 16 | n. 3 | p. 1-22 | ISSN 2317-2622. http://dx.doi.org/10.5935/2317-2622/direitomackenzie.v16n315705

RESUMO: Lidando com técnicas opacas de aprendizagem de máquinas, a questão crucial tornou-se a interpretabilidade do trabalho dos algoritmos e dos seus resultados. O artigo argumenta que a mudança para a interpretação requer um movimento da inteligência artificial para uma forma inovadora de comunicação artificial. Em muitos casos, o objetivo da explicação não é revelar os procedimentos das máquinas, e sim comunicar-se com elas e obter informação relevante e controlada. Como as explicações humanas não exigem transparência das ligações neurais ou processos de pensamento, as explicações algorítmicas não têm de revelar as operações da máquina, mas têm de produzir reformulações que façam sentido para os seus interlocutores. Esse movimento tem consequências importantes para a comunicação jurídica, em que a ambiguidade desempenha um papel fundamental. O problema da interpretação nos argumentos jurídicos, discute o artigo, não é que os algoritmos não explicam o suficiente, mas que devem explicar muito e com muita precisão, restringindo a liberdade de interpretação e a contestabilidade das decisões jurídicas. A consequência pode ser uma possível limitação da autonomia da comunicação jurídica que está na base do Estado de Direito moderno.

PALAVRAS-CHAVE: Explicação; interpretação; ambiguidade; Estado de Direito; inteligência artificial; IAE.

TRANSPARENCY VERSUS EXPLANATION: THE ROLE OF AMBIGUITY IN LEGAL AI

ABSTRACT: Dealing with opaque machine learning techniques, the crucial question has become the interpretability of the work of algorithms and their results. The paper argues that the shift towards interpretation requires a move from artificial intelligence to an innovative form of artificial communication. In many cases the goal of explanation is not to reveal the procedures of the machines but to communicate with them and obtain relevant and controlled information. As human explanations do not require transparency of neural connections or thought processes, so algorithmic explanations do not have to disclose the operations of the machine but have to produce reformulations that make sense to their interlocutors. This move has important consequences for legal communication,

where ambiguity plays a fundamental role. The problem of interpretation in legal arguments, the paper argues, is not that algorithms do not explain enough but that they must explain too much and too precisely, constraining freedom of interpretation and the contestability of legal decisions. The consequence might be a possible limitation of the autonomy of legal communication that underpins the modern rule of law.

KEYWORDS: Explanation; interpretation; ambiguity; rule of law; artificial intelligence; XAI.

1. Introdução: da inteligência artificial à comunicação artificial

Depois de repetidos "invernos" (RUSSELL; NORVIG, 2003, p. 29. CARDON; COINTET; MAZIERES, 2018, p. 173), a investigação IA parece estar agora numa nova "primavera" – na qual, no entanto, as máquinas, a forma de trabalhar e mesmo os problemas mudaram. Hoje falamos mais de algoritmos do que de computadores. Tomamos como certa a referência à web (incluindo a participação ativa dos usuários) e o fato de que os dados a serem processados não são escassos, e sim excessivamente abundantes. Estamos no mundo dos algoritmos de autoaprendizagem e de *big data*. Nessa nova fase, o problema central não é a capacidade ou o poder processante dos computadores[1]. Atualmente, a questão crucial é a interpretação, ou melhor, a interpretabilidade dos algoritmos[2] (DEANGELIS, 2014) e dos resultados do seu trabalho.

O artigo argumenta que a mudança para a interpretação exige que a investigação sobre o processamento de informação digital passe da referência à inteligência (artificial) para a referência a uma forma inovadora de comunicação, que pode ser definida como artificial (ESPOSITO, 2017, p. 249; ESPOSITO, 2021). O objetivo não é construir máquinas inteligentes, mas ser capaz de se comunicar com algoritmos para obter informação relevante e controlada. O que deve ser compreendido é a informação gerada nessa comunicação e não os processos das máquinas, que são e muitas vezes devem

1 O que as pessoas tentaram prever com a Lei de Moore e suas variantes.

2 Em discurso recente sobre a IA e suas transformações, o uso do termo 'algoritmo' é frequentemente impreciso. Naturalmente, a programação de computadores tem usado algoritmos desde o início e o termo já existia antes da cibernética. Neste texto, sigo o uso atual, por mais imperfeito que seja, e uso 'algoritmos' para me referir a técnicas avançadas de programação que utilizam aprendizagem de máquinas e grandes dados.

permanecer obscuros. Faço minhas observações nas próximas duas seções do artigo que tratam da questão da transparência e do objetivo das explicações.

A mudança da inteligência para a comunicação traz problemas e oportunidades em muitos campos diferentes, incluindo a complexa área de interpretação jurídica, abordada na seção 'Razão artificial e jurisprudência mecânica'. Lá, discuto o papel da interpretação para a autonomia do sistema jurídico, e, na seção seguinte, exploro a necessidade de ambiguidade na argumentação jurídica e os desafios resultantes para a utilização de algoritmos. A 'jurisprudência mecânica' pode afetar a prática jurídica e os princípios em que esta se baseia, notadamente o Estado de Direito.

2. A interpretação de máquinas incompreensíveis

A recente ênfase no problema da interpretação é uma consequência da inovação nas técnicas de programação e gestão de dados. Com métodos de aprendizagem profunda, e utilizando *big data*, os algoritmos aprendem de forma autônoma a executar as suas tarefas de formas não necessariamente previstas por seus programadores e que, em alguns casos, são incompreensíveis para os humanos, incluindo aqueles que os projetaram. Mesmo os programadores podem não compreender como a máquina procede e como ela alcança os seus resultados (GOODFELLOW; BENGIO; COURVILLE, 2016; BURRELL, 2016; WEINBERGER, 2017; GILPIN; BAU; YUAN; BAJWA; SPECTER; KAGAL, 2018. BUSUIOC, 2020). Quando se precisa entender os resultados e procedimentos dos algoritmos, é necessário interpretá-los, e não está claro como isso deve ser alcançado.

Algoritmos que trabalham com a aprendizagem de máquinas e *big data* estão ficando cada vez melhores em fazer cada vez mais coisas: eles produzem informação de forma rápida e precisa; eles estão aprendendo a dirigir carros com mais segurança e confiabilidade do que os humanos; eles podem responder às nossas perguntas, conversar, compor música e ler livros; e eles podem até mesmo escrever textos interessantes, apropriados, e – se necessário – engraçados. Eles alcançaram esses resultados, que parecem sugerir que as máquinas finalmente se tornaram inteligentes, já que seus programadores desistiram mais ou menos explicitamente de tentar reproduzir artificialmente os processos da inteligência humana. Os algoritmos funcionam de uma forma radicalmente diferente, que pode ser incompreensível para a nossa inteligência. A transparência, ou a falta dela, é, portanto, um problema.

Os algoritmos de aprendizagem de máquinas são de difícil entendimento, antes de tudo porque funcionam sem compreender seus materiais – eles fazem algo diferente. Programas recentes de tradução, por exemplo, não tentam entender os documentos que traduzem e os seus designers não confiam em nenhuma teoria de linguagem (BOELLSTORFF, 2013). Os algoritmos traduzem textos em mandarim sem conhecer mandarim; seus programadores também não conhecem. Os exemplos se multiplicam em todas as áreas nas quais os algoritmos têm mais sucesso, por exemplo, competindo com jogadores humanos no xadrez, pôquer e go (SILVER; HASSABIS, 2016), produzindo textos, programas de recomendação (PREY, 2018), reconhecimento de imagem e muitos outros. Os algoritmos não compreendem nada dos materiais com os quais estão lidando; eles "não raciocinam como as pessoas para escrever [ou, pode-se acrescentar, para trabalhar em geral] como as pessoas" (HAMMOND, 2015). Portanto, as operações das máquinas e seus resultados são muitas vezes obscuros para os observadores humanos.

Mesmo que sejam muito eficazes, porém, a confiança nas caixas pretas não é tranquilizadora, especialmente quando sabemos que as suas operações não são imunes a vieses e erros de vários tipos (PASQUALE, 2015). Em muitos casos, queremos verificar a correção dos resultados produzidos pelas máquinas, que podem ser errados ou inadequados de muitas maneiras diferentes, e com consequências diferentes. No campo médico, por exemplo, existe a preocupação de que os algoritmos possam não levar adequadamente em conta informações que, embora relevantes, podem não ser explícitas (HOLZINGER; LANGS; DENK; ZATLOUKAL; MÜLLER, 2019). Por exemplo, Caruana *et al.* discutem um algoritmo que previu que os pacientes asmáticos estavam com menor risco de morte por pneumonia, ignorando o fato de que os pacientes já vinham recebendo assistência médica intensa (CARUANA; LOU; GEHRKE; KOCH; STURM; ELHADAD, 2015). Em outros campos, como o policiamento (LUM; ISAAC, 2016), a concessão de crédito ao consumidor (O'NEIL, 2016), ou processos de admissão universitária (HAO, 2020), existe a preocupação de que, por meio de vieses sistêmicos ou de confirmação, eles possam reproduzir ou intensificar os desequilíbrios nos dados. Consequentemente, deseja-se poder verificar seus resultados e controlar a forma como são obtidos. No campo jurídico, discutido detalhadamente mais adiante, a obscuridade dos procedimentos algorítmicos pode comprometer a contestabilidade das decisões.

O recente ramo de pesquisa sobre '*explainable AI*' (XAI) tenta responder a essa preocupação desenvolvendo procedimentos para explicar as operações de algoritmos de autoaprendizagem (WACHTER; MITTELSTADT; FLORIDI, 2017; DOSHI-VELEZ;

KORTZ; BUDISH; BAVITZ; GERSHMAN; O´BRIEN; SCOTT; SCHIEBER; WALDO; WEINBERGER; WELLER; WOOD, 2017; MILLER, 2019). Os resultados esclarecem vários aspectos dos processos de interação com máquinas e são muitas vezes bastante úteis no gerenciamento de tais processos em situações específicas. No entanto, no caso de algoritmos de aprendizagem profunda, existe um obstáculo básico: se por explicação se entende um procedimento que permite aos observadores humanos compreender o que a máquina faz e por que a empresa não tem esperança. Os processos de algoritmos recentes que parecem inteligentes são intrinsecamente incompreensíveis para a inteligência humana. Como Weinberger afirma, exigir uma explicação nesse sentido equivaleria a "forçar a IA ser artificialmente estúpida o suficiente para que possamos compreender como ela chega à sua conclusão" (WEINBERGER, 2017; DOSHI-VELEZ; KIM, 2017; MONTAVON; SAMEK; MÜLLER, 2018; MONROE, 2018; RUDIN, 2019; BUSUIOC, 2020).

A estratégia deve ser diferente e, de fato, muitos projetos sobre XAI adotaram recentemente outra abordagem, compatível com a obscuridade radical dos processos algorítmicos (ROHLFING; CIMIANO; SCHARLAU; MATZNER; BUHL; BUSCHMEIER; ESPOSITO; GRIMMINGER; HAMMER; KERN; KOPP; THOMMES; NGOMO; SCHULTE; WACHSMUTH; WAGNER; WREDE, 2020). A noção chave é a transparência, frequentemente tomada como o primeiro elemento de projetos de IA explicáveis (ROSCHER; BOHN; DUARTE; GARCKE, 2020). Contudo, o debate envolve muitas outras noções relacionadas, cujas relações nem sempre são claras (MONROE, 2018; ANANNY; CRAWFORD, 2018; LIPTON, 2018; O'HARA, 2020), bem como as interações entre humano-computador muito além das questões de aprendizagem profunda que a desencadearam. Quando e por que se torna necessário explicar as operações dos algoritmos? O objetivo da explicação deve ser a transparência? Qual é a relação entre transparência e opacidade, e entre explicação e interpretação? O que deve ser explicado, a quem e para qual finalidade? E quando se pode dizer que uma explicação foi realmente produzida? A resposta a essas perguntas diz respeito à própria interpretação da IA e à sua relevância social.

3. Será que a explicação requer transparência?

No estudo sociológico da tecnologia, a falta de transparência tem sido um problema antigo (WEYER; SCHULZ-SCHAEFFER, 2009; LUHMANN, 2017). O problema se torna

ainda mais agudo no caso dos algoritmos. Aqui quero distinguir um tipo específico de não transparência, que pode ser chamado de opacidade, em relação a métodos recentes de aprendizagem de máquinas, tais como redes neurais, que usam algoritmos de "caixa preta" (BUHRMESTER; MÜNCH; ARENS, 2019). Os modelos correspondentes podem ser radicalmente incompreensíveis para os observadores humanos, por mais experientes que sejam. Outros modelos que são em princípio compreensíveis (não opacos), como algoritmos de "caixa branca" baseados em árvores de decisão (QUINLAN, 1986) ou programação de lógica indutiva (MUGGLETON; DE RAEDT, 1994), podem, no entanto, também se revelar não transparentes, devido ao seu tamanho ou complexidade, bem como pelo acesso restrito à informação relevante (como a obtenção e utilização de dados de formação ou o desenvolvimento e implementação do modelo), ou, em geral, porque o observador não tem as competências necessárias.

Na utilização de algoritmos, a não transparência é muito mais ampla do que a opacidade, e mesmo que fosse obrigatório que todas as fontes de dados e todos os procedimentos fossem acessíveis aos utilizadores, a maioria dos sistemas continuaria a ser incompreensível para seus usuários. Contudo, por si só, isso não é novo nem problemático: o funcionamento interno da tecnologia sempre foi incompreensível para a maioria dos usuários (LATOUR, 1999). A questão é que hoje os algoritmos fazem algo sem precedentes, diferente de outros sistemas tecnológicos: eles tomam decisões – sobre diagnósticos médicos, a seleção dos estudantes a serem admitidos nas universidades, as mudanças a serem feitas no go, as pessoas a receberem crédito ou liberdade condicional. São essas decisões que devem ser explicadas, e não os processos internos das máquinas. O objetivo da XAI é, na verdade, a explicação, não a transparência, e desse ponto de vista a opacidade dos sistemas de aprendizagem profunda não faz diferença; de qualquer forma, compreender a IA não é o problema.

O objetivo não é revelar os procedimentos das máquinas, mas, sim, fazer com que as próprias máquinas forneçam explicações que sejam informativas para o usuário. Não se pede que as máquinas sejam transparentes para os observadores humanos, mas que expliquem as suas decisões de forma que faça sentido para os seus interlocutores. E como seus interlocutores são sempre diferentes e localizados em situações e contextos diferentes, com interesses e necessidades diferentes, as explicações terão que ser diversas e específicas. A questão é fornecer explicações apropriadas aos diferentes usuários.

Isso é o que acontece quando os seres humanos tomam decisões, para as quais também podemos ser obrigados a oferecer explicações, dando pistas que permitam ao

destinatário dar sentido à decisão. Quando se obtém uma explicação, se obtém informação sobre a decisão sem ser informado sobre os processos neurofisiológicos ou psíquicos do explicador, os quais (felizmente) podem permanecer obscuros ou privados. Explicar as nossas decisões não requer a divulgação do nosso processo de pensamento, muito menos as conexões dos nossos neurônios. As explicações, afirma Luhmann (1990), são "reformulações com o benefício adicional de uma melhor conectividade". O emissor produz uma nova comunicação que fornece elementos adicionais relacionados ao pedido específico do interlocutor e suas necessidades. Em todo caso, esse é um processo inteiramente comunicativo: não precisamos acessar o cérebro ou a mente dos nossos interlocutores, nem precisamos acessar o mundo externo. Precisamos apenas obter pistas que permitam que a comunicação prossiga de uma forma controlada e não arbitrária.

A mesma abordagem pode ser prevista para lidar com os dilemas de explicação na interação com as máquinas de autoaprendizagem. Muitos têm sugerido que somente modelos inerentemente compreensíveis devem ser usados nos casos em que a explicação possa ser necessária (ROBBINS, 2019). Contudo, isso não resolve o problema geral do qual surge a necessidade de explicação[3]. Em vez disso, as máquinas, opacas ou não, deveriam ser capazes de produzir "reformulações" dos seus processos que correspondam às solicitações dos seus interlocutores e lhes permitam exercer a forma de controle adequada ao contexto. O desafio técnico nas interações com um parceiro digital é reproduzir a situação comunicativa em que as explicações são solicitadas e fornecidas entre seres humanos.

De fato, muitos projetos XAI recentes não tentam imitar os cálculos feitos pelo algoritmo, e sim produzir "explicações *post-hoc*" que reproduzem o que os seres humanos fazem na comunicação. A transparência não pode ser a solução, porque, como Lipton afirma, por mais que a transparência seja entendida (no nível de todo o modelo, no nível dos componentes individuais ou no nível dos algoritmos de treinamento), as explicações humanas não exibem transparência (LIPTON, 2018, p. 15). Os processos pelos quais as pessoas explicam suas decisões são distintos daqueles pelos quais as tomam e geralmente são produzidos após o fato, o que afeta a tomada de decisões. Do mesmo modo, no campo da XAI, os *designers* são programas de treinamento para produzir

3 "Se o ML está sendo usado para uma decisão que requer uma explicação, então ele deve ser explicável IA e um humano deve ser capaz de verificar se as considerações usadas são aceitáveis, mas se já sabemos quais considerações devem ser usadas para uma decisão, então não precisamos do ML" (ROBBINS, 2019).

explicações que ilustram (poderíamos dizer "reformular") o fato após o funcionamento dos algoritmos, sem impactar sua *performance*. Assim como os processos linguísticos que geram explicações humanas diferem dos processos neurais que produzem as decisões a serem explicadas, os processos que produzem explicações de modelos de IA também serão diferentes dos processos do modelo[4]. Eles podem, por exemplo, utilizar explicações verbais produzidas pela máquina, visualizações e explicações locais, como mapas de saliência (LIPTON, 2018, p. 15 *et seq*). A compreensão do usuário das explicações produzidas pela máquina não tem que se relacionar com os processos da máquina.

Essa perspectiva promissora implica uma mudança profunda em relação à abordagem que tem guiado os projetos de IA desde seu início nos anos 1950 – como o próprio nome Inteligência Artificial (IA) indica. De forma contraditória, os recentes projetos de XAI não estão focados na inteligência da máquina. Antes, o objetivo é o de produzir uma condição de ‹diálogo› entre o algoritmo e o usuário no qual a máquina fornece respostas, tomando como *input* os sempre diferentes pedidos de esclarecimento dos seus interlocutores (CIMIANO; RUDOLPH; HARTFIEL, 2010), e é capaz de participar numa metacomunicação (BATESON, 1972; LUHMANN, 1997, p. 250-251) que pode ter como objeto os processos da máquina ou os dados utilizados. O objetivo não é, e não pode ser, que os interlocutores compreendam esses processos, mas que interpretem o que a máquina comunica sobre esses processos de tal modo que possam exercer uma forma de controle. O debate sobre a explicação implica uma mudança da inteligência para a própria característica que permite que os algoritmos contribuam efetivamente para a produção de novas informações na nossa sociedade: a sua capacidade de participar na comunicação. As máquinas devem ser capazes de produzir explicações adequadas em resposta a diferentes pedidos dos seus interlocutores.

4. Razão artificial e jurisprudência mecânica

Se XAI implica um movimento do foco na inteligência para o foco na comunicação, a tarefa de observação sociológica seria mostrar como as interações com algoritmos afetam a comunicação na sociedade em geral (LUHMANN, 1993, p. 304; ESPOSITO,

4 Como as explicações bem-sucedidas por algoritmos não requerem acesso ao funcionamento dos algoritmos, a natureza de caixa preta dos algoritmos de aprendizado profundo não faz diferença para sua explicabilidade. Pelo contrário, algoritmos complexos como as redes neurais profundas podem ser mais eficientes no aprendizado, sendo as representações mais eficazes na comunicação com os usuários (LIPTON, 2018).

2017), e especificamente como as explicações algorítmicas funcionam como processos de comunicação que dependem da opacidade. Isso pode acontecer de diferentes maneiras em diferentes domínios da sociedade. Na pesquisa científica, por exemplo, na medicina, a atenção será direcionada para a possibilidade de descobrir estruturas causais nos dados[5]; no policiamento, será direcionada para a confiança nas decisões dos algoritmos; quando os algoritmos decidirem sobre a seleção de candidatos ou devedores, a questão será se as decisões algorítmicas estão em conformidade com os princípios éticos. Essa seção explora o campo jurídico: como a falta de transparência e sua gestão no funcionamento dos algoritmos pode afetar a prática jurídica e seus pressupostos, notadamente o Estado de Direito.

No campo jurídico, hoje os algoritmos são capazes de cumprir muitas tarefas de forma barata, eficaz e rápida: eles podem automatizar o preenchimento de documentos, realizar a *due diligence*, reunir e analisar dados antigos, classificar por meio de informações jurídicas e realizar outras atividades que anteriormente exigiam trabalho humano. As oportunidades resultantes e os riscos associados ao trabalho provocaram um amplo debate tanto no campo jurídico quanto em outros setores (SUSSKIND, 2008). A questão que queremos abordar aqui é mais abstrata e complexa, envolvendo o papel da interpretação nos argumentos jurídicos. Aqui, também, os computadores podem ser utilizados de forma útil para realizar muitas tarefas. As pessoas falam de "jurisprudência mecânica" (WALTON; MACAGNO; SARTOR, 2021) ou "ciência jurídica computacional" (LETTIERI; ALTAMURA; GIUGNO; GUARINO; MALANDRINO; PULVIRENTI; VICIDOMINI; ZACCAGNINO, 2018), sistemas computacionais de raciocínio jurídico capazes de explorar bases de dados jurídicos (ALETRAS; TSARAPATSANIS; PREOŢIUC-PIETRO; LAMPOS, 2016), identificar regras relevantes, tomar decisões (BINNS, 2020), gerar argumentos e, também, explicar sua cadeia de raciocínio aos usuários (ASHLEY, 2017). As máquinas participam de forma autônoma da comunicação jurídica: elas podem gerar informações juridicamente relevantes, elaborar um argumento e até mesmo explicá-lo.

O problema é mais profundo e não diz respeito apenas à possível ameaça às habilidades dos trabalhadores humanos e seus empregos. Diz respeito aos fundamentos do direito positivo moderno, que envolvem a autonomia do direito e a questão da

5 O animado debate sobre a diferença entre correlação e causalidade na ciência é um caso influente, desencadeando um profundo repensar de questões epistemológicas básicas, como a relação entre explicações e previsões (PEARL, 2000; PEARL; MACKENZIE, 2018; BREIMAN, 2001; SHMUELI, 2020; SOBER, 2016).

interpretação. Como Hildebrandt argumentou, nossa forma de sistema jurídico se desenvolveu como resultado da disseminação da máquina de impressão e das mudanças resultantes na forma como produzimos, escrevemos e lemos textos (HILDEBRANDT, 2020; LUHMANN, 1993, p. 349). A máquina de impressão produz textos padronizados, idênticos e imutáveis, que são retirados da "*mouvance*" da comunicação oral e dos manuscritos (ZUMTHOR, 1972; EISENSTEIN, 1979) – livros que escapam à prática do comentário. Em textos anteriores, numa cultura que se manteve preponderantemente oral, foram adicionados glosas e comentários em cada leitura e se tornaram parte do texto, que mudava ("*moved*") continuamente, produzindo cada vez uma comunicação diferente (ASSMANN; GLADIGOW, 1995). O texto "móvel" incorporou a interpretação.

Quando, com a máquina de impressão, o texto se tornou fixo e permaneceu o mesmo em todas as leituras, as interpretações se multiplicaram e se tornaram variáveis. A escrita, argumenta Luhmann, dá origem à diferença entre texto e interpretação, que a máquina de impressão generaliza (LUHMANN, 1993, p. 362). O texto fixo deve ser interpretado para fazer sentido no contexto específico. Entretanto, as situações em que um texto é lido são todas únicas, diferentes de qualquer outra; se o texto permanecer o mesmo, a forma de considerá-lo deve mudar. A pluralidade de interpretações é inevitável e legítima: como os contextos e as circunstâncias são sempre diferentes, as interpretações devem variar para levá-los em conta (ESPOSITO, 2002, p. 226-227). As interpretações de um mesmo texto, portanto, podem ser sempre diferentes, e qualquer interpretação pode ser contestada.

Isso acontece em todos os campos que têm a ver com textos, mas na prática jurídica assume uma forma mais complicada[6]. Se as leis são textos escritos e as decisões judiciais também assumem essa forma, é necessário muito trabalho de interpretação para levar em conta a variedade de circunstâncias e casos jurídicos. Os juízes interpretam as leis e casos anteriores, e os seus observadores (advogados, litigantes, público) interpretam as suas decisões. De acordo com Hildebrandt (2020), a liberdade de interpretação é a base do Estado de Direito moderno. Essa liberdade é a base da autonomia do judiciário. Ela permite ao judiciário seguir a própria lógica e critérios. Estes não são ditados pela soberania e podem entrar em conflito com os princípios e as preferências do poder político. Nos termos de Fried, "a racionalidade da lei é uma racionalidade à parte",

6 Sobre a performatividade da linguagem, ver Austin, 1962. No campo jurídico, esta é uma condição básica: as palavras pronunciadas por um juiz ou um legislador são fatos imediatos e têm consequências concretas.

que não segue os princípios da racionalidade geral, mas apenas a "razão artificial da lei" (FRIED, 1981, p. 35, 39 e 58)[7].

A autonomia de interpretação é um requisito básico para a independência da lei, mas não significa arbitrariedade ou obscuridade. As decisões dos juízes devem ser explicadas, ou seja, motivadas (em termos jurídicos) de acordo com a racionalidade específica da lei, explicitando as razões em que se baseiam. De acordo com essa racionalidade, então, as explicações são interpretadas e as decisões podem ser contestadas. "O propósito da interpretação não é assegurar que todos os leitores compreendam o texto da mesma forma, mas que diferentes pessoas que enfrentam o mesmo texto participem de uma comunicação unitária" (LUHMANN, 1993, p. 362). Esse é o tipo de transparência exigida pelo funcionamento controlado do sistema jurídico e aquele segundo o qual a possível transparência dos algoritmos deve ser avaliada. A explicação dada pela inteligência artificial na jurisprudência mecânica atende os requisitos da "razão artificial da lei"? Uma decisão tomada com base em procedimentos automatizados pode ser justificada de modo a permitir o funcionamento da comunicação legal e possivelmente a contestação pelas pessoas envolvidas? A falta de transparência dos algoritmos, que, como vimos, é inevitável em seu uso comunicativo, é compatível com as exigências de transparência das decisões legais?

5. O papel da ambiguidade nos argumentos jurídicos

Em um primeiro nível, esse parece ser o caso. Que os processos digitais que conduzem à decisão são diferentes daqueles de nossa inteligência e possivelmente não são acessíveis ou compreensíveis para os observadores humanos, mas, no que diz respeito à comunicação legal, isso não marca necessariamente uma cesura com as decisões tomadas pelos agentes humanos. Como afirmam Canale e Tuzet, "motivação jurisdicional não consiste no relato psicológico do processo que conduziu à decisão, mas na indicação das razões legais que a justificam" (CANALE; TUZET, 2020), ou, como assevera Luhmann, "o argumento não reflete o que o leitor tem em mente" (LUHMANN, 1993, p. 362). Uma motivação correta não implica que os pensamentos e passos que levaram à decisão

7 A Teoria sociológica dos Sistemas descreve essas condições como *out-differentiation* (*Ausdifferenzierung*) do sistema jurídico na sociedade moderna. Ver Luhmann (n. 33), p. 743 *et seq*.

sejam descritos e, portanto, pode ser argumentado, nem deve ser necessário descrever os processos seguidos pelo algoritmo para chegar ao seu resultado. Não é necessariamente um problema que os processos digitais sejam incompreensíveis para os seres humanos, se o algoritmo for capaz de explicar sua decisão num sentido comunicativo, ou seja, de indicar de forma compreensível as razões legais que a levaram ou, no sentido de Fried (1981), a razão artificial em que se baseia.

Em um segundo nível, no entanto, as coisas são mais complicadas. De uma perspectiva sociológica, o desempenho da lei para a sociedade como um todo é a "absorção da incerteza" na gestão do litígio (LUHMANN, 1966, p. 56-57)[8]. Deve ser possível confiar no fato de que as regras legais são aplicadas a casos concretos e de uma forma válida (LETTIERI, 2020, p. 72). Para absorver a incerteza, a validade deve ser argumentada (motivada), ou seja, a decisão legal deve ser justificada, fornecendo fundamentos para ela. Como os casos a serem tratados são sempre diferentes, os fundamentos devem ser apropriados ao contexto (WALTON; MACAGNO; SARTOR, 2021), mas a própria decisão sobre o que conta como contexto pode ser controversa e levar a dúvidas e discordâncias (EASTERBROOK, 2017, p. 81, 83-84). Na maioria dos casos, além disso, muitas das provas apresentadas por ambas as partes para apoiar seus argumentos são baseadas em regras e precedentes conflitantes (BERMAN; HAFNER, 1988). Embora todas as decisões jurídicas se refiram ao mesmo conjunto de regras, os argumentos (explicações) devem ser diferentes, caso a caso, e coordenados de forma flexível entre si.

Para que a coordenação seja possível, a ambiguidade desempenha um papel fundamental na comunicação jurídica (LETTIERI, 2020. HILDEBRANDT, 2020. HOFFMANN-RIEM, 2020). Os argumentos "são tipicamente vagos e ambíguos" (WALTON; MACAGNO; SARTOR, 2021, p. 4), ou seja, "suscetíveis de mais de uma interpretação razoável" (SOLAN, 2004). As normas legais são caracterizadas por múltiplas camadas de ambiguidade, que dificultam sua organização em um todo formal e totalmente consistente. Em casos típicos de argumentos jurídicos "a inconsistência é a norma" (WALTON; MACAGNO; SARTOR, 2021, p. 5; MATTARELLA, 2011), e, na verdade, o objetivo do argumento só pode ser "evitar inconsistências visíveis" (LUHMANN, 1993, p. 356). O objetivo real do argumento não é conseguir coerência lógica abstrata, mas fazer com que os fundamentos da decisão pareçam convincentes – e uma justificativa

8 Na clássica definição de March e Simon: "A absorção da incerteza ocorre quando as inferências são retiradas de um conjunto de evidências e as inferências, ao invés das próprias evidências, são, então, comunicadas" (MARCH; SIMON, 1958, p. 165).

jurídica é convincente não porque todos os seus passos foram verificados: "A racionalidade da gestão de problemas jurídicos reside... não na exatidão lógica de suas conclusões... Deve bastar que convença a todos de que convenceu seu autor" (LUHMANN, 1966, p. 55, 59). A motivação (explicação) parece convincente quando todos estão convencidos de que outros a acham convincente. A eficácia retórica conta mais do que a consequência lógica das etapas do argumento, que não é examinada em detalhes.

Advogados e juízes, que são "os mestres da razão artificial da lei", são por experiência e expertise profissional muito competentes para lidar com a ambiguidade e usá-la para fins retóricos, por exemplo, aplicando "uma intuição treinada e disciplinada em que a multiplicidade de detalhes é muito extensa para permitir que nossas mentes trabalhem sobre ela dedutivamente" (FRIED, 1981, p. 57). A tarefa dos advogados, afirma Garfinkel, é tornar ambíguas as interpretações de fatos e leis (GARFINKEL, 1967, p. 111). Funciona bem quando se interage com seres humanos, pois para uma comunicação eficaz é suficiente regular "a apresentação, não a produção da decisão" (LUHMANN, 1966, p. 106). Advogados e juízes devem apresentar um relato convincente das decisões que tomam, mas a sua interpretação pode e muitas vezes deve permanecer vaga, pois "não se preocupa com a forma como compreendemos ou produzimos textos, mas sim com a forma como estabelecemos a aceitabilidade de uma leitura específica dos mesmos" (WALTON; MACAGNO; SARTOR, 2021, p. 9). O que os observadores interpretam é a interpretação geralmente ambígua por parte do juiz ou do advogado.

Para os algoritmos, contudo, a ambiguidade é um desafio. A gestão competente da vagueza é notoriamente um problema para as máquinas, que vem sendo discutido há décadas nos discursos sobre os limites da inteligência artificial (DREYFUS, 1972). Ainda hoje é difícil, para os algoritmos, lidar com os vários níveis de ambiguidade sempre presentes na comunicação humana ou, no campo jurídico, gerenciar a multiplicidade de possíveis interpretações de regras e normas (LETTIERI, 2020). Além disso, se o foco passa da inteligência das máquinas (o que elas podem compreender e como) para sua participação na comunicação, surgem outros problemas relacionados à ambiguidade: não só a dificuldade das máquinas em lidar com a ambiguidade da comunicação humana, mas também a dificuldade de elas mesmas gerarem uma comunicação ambígua, ou seja, de administrarem de forma competente a ambiguidade exigida pelos argumentos jurídicos.

As explicações jurídicas produzidas pelos algoritmos devem ser elas mesmas ambíguas, assim como são aquelas que resultam da interpretação de normas jurídicas

por humanos. A ambiguidade não é, como tendemos a pensar, oposta à transparência (ANANNY; CRAWFORD, 2018; OLSEN, 2014; HEIMSTÄDT; DOBUSCH, 2020), mas, ao contrário, é necessária para fornecer a multiplicidade de interpretações jurídicas indispensável para a contestabilidade. Como Hildebrandt afirma, "devido à ambiguidade inerente à linguagem humana, os ICIs[9] orientados por texto geram um tipo específico de multi-interpretabilidade que, por sua vez, gera um tipo específico de contestabilidade" (HILDEBRANDT, 2020, p. 7-8). Para contestar uma decisão, é preciso ser capaz de desenvolver uma perspectiva sobre a decisão que seja independente daquela fornecida pelo tomador da decisão (O'HARA, 2020), ou seja, questionar a sua interpretação. Contudo, para fazer isso, a motivação deve parecer juridicamente ambígua – isto é, deve ser, como vimos, suscetível a mais de uma interpretação razoável. A máquina que não tem a própria perspectiva não interpreta, portanto, suas explicações carecem de ambiguidade. As explicações que ela oferece são reformulações das decisões que são tomadas seguindo outras regras, portanto não faz sentido perguntar o que o algoritmo significou – os algoritmos não significam nada.

A falta de uma gestão competente da ambiguidade é um problema que também é percebido em experiências que tentam realizar uma forma de XAI no campo jurídico. Mesmo para os modelos computacionais mais recentes que produzem argumentação jurídica, a falta de ambiguidade é uma restrição (WALTON; MACAGNO; SARTOR, 2021, p. 11) muito além do que é exigido na comunicação legal entre seres humanos guiados pelo imperativo de parecer convincente e absorver a incerteza. Paradoxalmente, então, pode-se dizer que o problema da interpretação na argumentação jurídica – mesmo e precisamente quando se trata de algoritmos que são obscuros para a inteligência humana – não é que a máquina não explique o suficiente, mas que deve explicar demais, e com muita precisão. Como reconhecem os estudiosos nesse campo, esse nível de detalhe pode obscurecer, em vez de iluminar, a prática da comunicação jurídica:

> Estamos bem cientes de que, ao utilizar a abordagem de argumentação estruturada e formalista, existe o perigo de confundir os leitores mais do que explicar-lhes como os tribunais podem fazer um melhor trabalho de luta com os difíceis (chamados de perversos) problemas de interpretação estatutária (WALTON; MACAGNO; SARTOR, 2021, p. 12)[10].

9 "Information and communication infrastructures". Tradução: Infraestruturas de informação e comunicação.

10 Original: "We are well aware that in using the structured and formalistic argumentation approach there is the danger of confusing readers more than explaining to them how the courts can do a better job of grappling with the hard (so-called wicked) problems of statutory interpretation".

Por um lado, portanto, há o risco de que a explicação não seja convincente. Por outro lado, se for convincente, talvez possa surgir um problema ainda mais grave: podem ser impostos limites à liberdade de interpretação que sustenta a autonomia da comunicação jurídica, e pode haver o risco de que o uso de modelos automatizados possa alterar características fundamentais do Estado de Direito (LETTIERI, 2020). Como vimos acima, a "razão artificial da lei" não coincide com a racionalidade geral da sociedade ou mesmo com a coerência abstrata de um argumento lógico. A jurisprudência mecânica, entretanto, quando identifica e aplica as regras jurídicas relevantes ao caso em questão, não funciona com os argumentos retóricos eficazes que caracterizam o raciocínio e a interpretação jurídica (ASHLEY, 2017), que são possivelmente ambíguos e não totalmente coerentes. A autonomia da comunicação jurídica, com todas as suas implicações na estrutura da sociedade moderna, pode tomar uma forma diferente como consequência da intervenção de algoritmos na comunicação.

Que liberdade resta para aqueles que devem interpretar um argumento jurídico "mecânico"? E, em particular, como a decisão pode ser contestada? Os argumentos produzidos pelos algoritmos não são interpretações, contingentes e passíveis de revisão, mas descrições de uma série de etapas formais. O observador pode descobrir um erro formal e contestar a decisão a esse nível. Entretanto, não pode explorar e contestar a interpretação, porque a máquina não interpretou nada. Todos os argumentos que se referem a razões e motivos de interpretação, a saber, "os fatores que podem levar um tomador de decisão a selecionar uma ou outra interpretação" (WALTON; MACAGNO; SARTOR, 2021, p. 97 *et seq.*), podem ser de fato desqualificados, e com eles um componente fundamental da comunicação jurídica na sociedade moderna.

6. Conclusão: comunicação com as máquinas

A observação do desafio colocado pelos algoritmos opacos sob a perspectiva da comunicação revela uma multiplicidade de perguntas fascinantes e difíceis. Algumas questões se dissolvem, como aquela baseada no teste *Turing*: interagimos rotineiramente com parceiros digitais sem nos perguntarmos se eles são seres humanos ou não. Outras questões tomam uma forma diferente, por exemplo, o complexo problema do viés, que envolve tanto a dimensão do viés algorítmico, refletindo os valores dos programadores (CRAWFORD, 2016), como a do viés de dados, dependendo da entrada descoordenada

de bilhões de participantes, sensores e outras fontes digitais (MEHRABI MORSTATTER; SAXENA; LERMAN; GALSTYAN, 2021). Ainda, outras questões surgem relacionadas à experiência prática acumulada em muitos campos. O uso de algoritmos para tarefas específicas está quase inadvertidamente levando ao surgimento de diversos, e extremamente complexos, problemas relacionados ao seu envolvimento na comunicação. A questão da interpretação na argumentação jurídica é um exemplo particularmente significativo. O problema não é como as máquinas funcionam, mas como elas participam da comunicação jurídica.

REFERÊNCIAS

ALETRAS, N.; TSARAPATSANIS, D.; PREOȚIUC-PIETRO, D.; LAMPOS, V. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing Perspective. *PeerJ Computer Science*, 2:e93, 2016. Disponível em: https://doi.org/10.7717/peerj-cs.93. Acesso em: 18 nov. 2022.

ANANNY, M.; CRAWFORD, K. Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. *New Media & Society*, v. 20, n. 3, p. 973-989, 2018. Disponível em: https://doi.org/10.1177/1461444816676645. Acesso em: 18 nov. 2022.

ASHLEY, K. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge: Cambridge University Press, 2017.

ASSMANN, J.; GLADIGOW, B. (eds.). *Text und Kommentar*. Archäologie der Literarischen Kommunikation IV. Leiden: Brill; Fink, 1995.

AUSTIN, J. L. *How to Do Things with Words*. Oxford: Oxford University Press, 1962.

BATESON, G. *Steps to an Ecology of Mind*. Chicago: University of Chicago Press, 1972.

BERMAN, D.; HAFNER, C. Obstacles to the Development of Logic-Based Models of Legal Reasoning. *In*: WALTER, C. (ed.). *Computer Power and Legal Language*. Westport: Greenwood Press, 1988.

BINNS, R. Analogies and disanalogies between machinedriven and human-driven legal judgement. *Journal of Cross-disciplinary Research in Computational Law*, v. 1, n. 1, p. 1-16, dez. 2020. Disponível em: https://journalcrcl.org/crcl/article/view/5. Acesso em: 21 nov. 2022.

BOELLSTORFF, T. Making Big Data, in Theory. *First Monday*, v. 18, n. 10, set. 2013. Disponível em: https://doi.org/10.5210/fm.v18i10.4869. Acesso em: 21 nov. 2022.

BREIMAN, L. Statistical Modeling: The Two Cultures. Statistical Science 199, v. 16, n. 3, p. 199-215, ago. 2001. Disponível em: https://www.jstor.org/stable/2676681. Acesso em: 21 nov. 2022.

BUHRMESTER, V.; MÜNCH, D.; ARENS, M. Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Computer Science*, arXiv:1911.12116, p. 1-22, 2019. Disponível em: https://doi.org/10.48550/arXiv.1911.12116. Acesso em: 21 nov. 2022.

BURRELL, J. How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, v. 3, n. 1, p. 1-12, 2016. Disponível em: https://doi.org/10.1177/2053951715622512. Acesso em: 21 nov. 2022.

BUSUIOC, M. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, v. 81, n. 5, p. 825-836, 2020. Disponível em: https://doi.org/10.1111/puar.13293. Acesso em: 21 nov. 2022.

CANALE, D.; TUZET, G. *La Giustificazione della Decisione Giudiziale*. Torino: Giappichelli, 2020.

CARDON, D.; COINTET J.-P.; MAZIERES, A. La revanche des neurons. L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux*, v. 211, n. 5, p. 173-220, 2018. Disponível em: https://doi.org/10.3917/res.211.0173. Acesso em: 21 nov. 2022.

CARUANA, R.; LOU, Y.; GEHRKE, J.; KOCH, P.; STURM, P.; ELHADAD, N. Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *In*: PROCEEDINGS OF ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 21, 2015, Sydney. *Proceedings* [...]. Sydney, Association for Computing Machinery, 2015. Disponível em: https://doi.org/10.1145/2783258.2788613. Acesso em: 21 nov. 2022.

CIMIANO, P.; RUDOLPH, S.; HARTFIEL, H. Computing Intensional Answers to Questions - An Inductive Logic Programming Approach. *Data & Knowledge Engineering*, v. 69, n. 3, p. 261-278, mar. 2010. Disponível em: https://doi.org/10.1016/j.datak.2009.10.008. Acesso em: 21 nov. 2022.

CRAWFORD, K. Artificial Intelligence's White Guy Problem. *The New York Times*, New York, 25 jun. 2016. Disponível em: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html. Acesso em: 21 nov. 2022.

DEANGELIS, S. F. Artificial Intelligence. How Algorithms Make Systems Smart. *Wired*, Boone, 2014. Disponível em: https://www.wired.com/insights/2014/09/artificial-intelligence-algorithms-2/. Acesso em: 23 jun. 2021.

DOSHI-VELEZ, F.; KIM, B. Towards A Rigorous Science of Interpretable Machine Learning. *Statistics*, arXiv:1702.08608v2, p. 1-13, 2017. Disponível em: https://doi.org/10.48550/arXiv.1702.08608. Acesso em: 21 nov. 2022.

DOSHI-VELEZ, F.; KORTZ, M.; BUDISH, R.; BAVITZ, C.; GERSHMAN, S.; O'BRIEN, D.; SCOTT, K., SCHIEBER, S.; WALDO, J., WEINBERGER, D.; WELLER, A.; WOOD, A. Accountability of AI Under the Law: The Role of Explanation. *Computer Science*, arXiv:1711.01134v3 , p. 1-21, 2017. Disponível em: https://doi.org/10.48550/arXiv.1711.01134. Acesso em: 21 nov. 2022.

DREYFUS, H. *What Computers Can't Do*. Cambridge: The MIT Press, 1972.

EASTERBROOK, F. H. The Absence of Method in Statutory Interpretation. *Chicago Law Review*, v. 84, n. 81, p. 81-97, 2017. Disponível em: https://lawreview.uchicago.edu/publication/absence-method-statutory-interpretation. Acesso em: 21 nov. 2022.

EISENSTEIN, E. L. *The Printing Press as an Agent of Change*. Communications and Cultural Transformations in Early-Modern Europe. Cambridge: Cambridge University Press, 1979.

ESPOSITO, E. *Soziales Vergessen*. Formen und Medien des Gedächtnisses der Gesellschaft. Berlin: Suhrkamp, 2002.

ESPOSITO, E. Artificial Communication? The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*, v. 46, n. 4, p. 249-265, ago. 2017. Disponível em: https://doi.org/10.1515/zfsoz-2017-1014. Acesso em: 21 nov. 2022.

ESPOSITO, E. *Artificial Communication*. How Algorithms Produce Social Intelligence. Cambridge: MIT Press, 2021.

FRIED, C. Artificial Reason of the Law or: What Lawyers Know. *Texas Law Review*, v. 60, n. 1, p. 23-32, 1981. Disponível em: https://informallogic.ca/index.php/informal_logic/article/view/2598/2039. Acesso em: 21 nov. 2022.

GARFINKEL, H. *Studies in Ethnomethodology*. Hoboken: Prentice Hall, 1967.

GILPIN, L. H.; BAU, D.; YUAN, B. Z.; BAJWA, A.; SPECTER, M.; KAGAL, L. *Explaining Explanations*: An Overview of Interpretability of Machine Learning. *Computer Science*, arXiv:1806.00069v3 , p. 1-10, 2018. Disponível em: https://doi.org/10.48550/ar-Xiv.1806.00069. Acesso em: 21 nov. 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning* (Adaptive Computation and Machine Learning). Cambridge: MIT Press, 2016.

HAMMOND, K. *Practical Artificial Intelligence for Dummies*. Hoboken: Wiley, 2015.

HAO, K. The UK exam debacle reminds us that algorithms can't fix broken systems. *MIT Technology Review*, Cambridge, 20 ago. 2020. Disponível em: https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/#:~:text=Tech%20policy-,The%20UK%20exam%20debacle%20reminds%20us%20that%20algorithms%20can't,for%20standardization%20above%20all%20else.&text=When%20the%20UK%20first%20set,the%20premise%20seemed%20perfectly%20reasonable. Acesso em: 2 out. 2020.

HEIMSTÄDT, M.; DOBUSCH, L. Transparency and Accountability: Causal, Critical and Constructive Perspectives. *Organization Theory*, v. 1, n. 4, p. 1-12, 2020. Disponível em: https://doi.org/10.1177/2631787720964216. Acesso em: 21 nov. 2022.

HILDEBRANDT, M. *Law for Computer Scientists and Other Folk*. Oxford: Oxford University Press, 2020.

HILDEBRANDT, M. The Adaptive Nature of Text-Driven Law. *Journal of Cross-disciplinary Research in Computational Law*, v. 1, n. 1, p. 1-15, nov. 2020. Disponível em: https://journalcrcl.org/crcl/article/view/2. Acesso em: 21 nov. 2022.

HOFFMANN-RIEM, W. Legal Technology/Computational Law: Preconditions, Opportunities and Risks. *Journal of Cross-disciplinary Research in Computational Law*, v. 1, n. 1, p. 1-16, nov. 2020. Disponível em: https://journalcrcl.org/crcl/article/view/7. Acesso em: 21 nov. 2022.

HOLZINGER A.; LANGS, G.; DENK, H.; ZATLOUKAL, K.; MÜLLER, H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, v. 9, n. 4, e1312, abr. 2019. Disponível em: https://doi.org/10.1002/widm.1312. Acesso em: 21 nov. 2022.

LATOUR, B. *Pandora's Hope*: Essays on the Reality of Science Studies. Harvard: Harvard University Press, 1999.

LETTIERI, N.; ALTAMURA, A.; GIUGNO, R.; GUARINO, A.; MALANDRINO, D.; PULVIRENTI, A.; VICIDOMINI, F.; ZACCAGNINO, R. Ex Machina: Analytical Platforms, Law and the Challenges of Computational Legal Science. *Future Internet*, v. 10, n. 5, p. 1-25, 2018. Disponível em: https://doi.org/10.3390/fi10050037. Acesso em: 21 nov. 2022.

LETTIERI, N. Law, Rights, and the Fallacy of Computation. *Jura Gentium*, v. XVII, n. 2, p. 72-87, 2020. Disponível em: https://www.juragentium.org/Centro_Jura_Gentium/la_Rivista_files/JG_2020_2/JG_2020_2_Lettieri.pdf. Acesso em: 21 nov. 2022.

LIPTON, Z. C. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *ACM Queue*, v. 16, n. 3, p. 31-57, 2018. Disponível em: https://doi.org/10.1145/3236386.3241340. Acesso em: 21 nov. 2022.

LUHMANN, N. *Die Gesellschaft der Gesellschaft*. Berlin: Suhrkamp, 1997.

LUHMANN, N. *Recht und Automation in der öffentlichen Verwaltung*. Berlin: Duncker & Humblot, 1966.

LUHMANN, N. *Die Wissenschaft der Gesellschaft*. Berlin: Suhrkamp, 1990.

LUHMANN, N. *Das Recht der Gesellschaft*. Berlin: Suhrkamp, 1993.

LUHMANN, N. *Die Kontrolle von Intransparenz*. Berlin: Suhrkamp, 2017.

LUM, K.; ISAAC, W. To Predict and Serve. *Significance*, v. 13, n. 5, p. 14-19, out. 2016. Disponível em: https://doi.org/10.1111/j.1740-9713.2016.00960.x. Acesso em: 21 nov. 2022.

MARCH, J. G.; SIMON, H. A. *Organizations*. Hoboken: Wiley, 1958.

MATTARELLA, B. G. *La trappola delle leggi*: molte, oscure, complicate. Bologna: Il Mulino 2011.

MEHRABI, N; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, v. 54, n. 6, p. 1-35, jul. 2021. Disponível em: https://doi.org/10.1145/ 3457607. Acesso em: 21 nov. 2022.

MILLER, T. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence*, v. 267, p. 1-38, fev. 2019. Disponível em: https://doi.org/10.1016/j.artint.2018.07.007. Acesso em: 21 nov. 2022.

MONROE, D. AI, Explain Yourself. *Communications of the ACM*, v. 61, n. 11, p. 11-13, nov. 2018. Disponível em: https://doi.org/10.1145/3276742. Acesso em: 21 nov. 2022.

MONTAVON, G.; SAMEK, W.; MÜLLER, K.-R. Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing*, v. 73, p. 1-15, fev. 2018. Disponível em: https://doi.org/10.1016/j.dsp.2017.10.011. Acesso em: 21 nov. 2022.

MUGGLETON, S.; DE RAEDT, L. Inductive Logic Programming: Theory and Methods. *The Journal of Logic Programming*, v. 19-20, p. 629-679, maio/jun.1994. Disponível em: https://doi.org/10.1016/0743-1066(94)90035-3. Acesso em: 21 nov. 2022.

O'HARA, K. Explainable AI and the Philosophy and Practice of Explanation. *Computer Law & Security Review*, v. 39, 105474, nov. 2020. Disponível em: https://doi.org/10.1016/j.clsr.2020.105474. Acesso em: 21 nov. 2022.

O'NEIL, C. *Weapons of Math Destruction*. New York: Crown Publishing Group, 2016.

OLSEN, J. P. Accountability and Ambiguity. In: Bovens, M.; Goodin, R. E.; Schillemans, T. (eds.). *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press, 2014.

PASQUALE, F. *The Black Box Society. The Secret Algorithms That Control Money and Information*. Harvard: Harvard University Press, 2015.

PEARL, J. *Causality*. Cambridge: Cambridge University Press, 2000.

PEARL, J.; MACKENZIE, D. *The Book of Why*: The New Science of Cause and Effect. New York: Basic Books, 2018.

PREY, R. Nothing Personal: Algorithmic Individuation on Music Streaming Platforms. *Media, Culture & Society*, v. 40, n. 7, p. 1086-1100, 2018. Disponível em: https://doi.org/10.1177/0163443717745147. Acesso em: 21 nov. 2022.

QUINLAN, J. R. Induction of Decision Trees. *Machine Learning*, v. 1, p. 81-106, 1986. Disponível em: https://doi.org/10.1007/BF00116251. Acesso em: 21 nov. 2022.

ROBBINS, S. A Misdirected Principle With a Catch: Explicability for AI. Minds and Machines, v. 29, p. 495-514, 2019. Disponível em: https://doi.org/10.1007/s11023-019-09509-3. Acesso em: 22 nov. 2022.

ROHLFING, K. J.; CIMIANO, P.; SCHARLAU, I.; MATZNER, T.; BUHL, H. M.; BUSCHMEIER, H.; ESPOSITO, E.; GRIMMINGER, A.; HAMMER, B.; KERN, F.; KOPP, S.; THOMMES, K.; NGOMO, A.-C. N.; SCHULTE, C.; WACHSMUTH, H.; WAGNER, P.; WREDE, B. Explanations as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. IEEE Transactions on Cognitive and Developmental Systems, v. 13, n. 3, p. 717-728, set. 2021. Disponível em: https://doi.org/10.1109/TCDS.2020.3044366. Acesso em: 22 nov. 2022.

ROSCHER, R.; BOHN, B.; DUARTE, M. F.; GARCKE, J. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access, v. 8, p. 42200-42216, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2020.2976199. Acesso em: 22 nov. 2022

RUDIN, C. Stop Explaining Black Box Machine Learning Models for High Stake Decisions and Use Interpretable Models Instead. Nature Machine Intelligence, v. 1, p. 206-215, maio 2019. Disponível em: https://doi.org/10.1038/s42256-019-0048-x. Acesso em: 22 nov. 2022.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. Hoboken: Prentice Hall, 2003.

SHMUELI, G. To Explain or to Predict? Statistical Science, v. 25, n. 3, p. 289-310, ago. 2020. Disponível em: https://doi.org/10.1214/10-STS330. Acesso em: 22 nov. 2022.

SILVER, D.; HASSABIS, D. AlphaGo: Mastering the Ancient Game of Go with Machine Learning. Google DeepMind, 27 jan. 2016. Disponível em: https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html. Acesso em: 23 jun. 2021.

SOBER, E. Ockham's razors: a user's manual. Cambridge: Cambridge University Press 2016.

SOLAN, L. Pernicious Ambiguity in Contracts and Statutes. Chicago-Kent Law Review, v. 79, n. 3, p. 859-888, 2004. Disponível em: https://scholarship.kentlaw.iit.edu/cklawreview/vol79/iss3/22. Acesso em: 22 nov. 2022.

SUSSKIND, R. The End of Lawyers? Rethinking the Nature of Legal Services. Oxford: Oxford University Press, 2008.

WACHTER, S.; MITTELSTADT, B.; FLORIDI, L. Transparent, Explainable and Accountable AI for Robotics. Science Robotics, v. 2, n. 6, eaan6080, maio 2017. Disponível em: https://doi.org/10.1126/scirobotics.aan6080. Acesso em: 22 nov. 2022.

WALTON, D.; MACAGNO, F.; SARTOR, G. Statutory Interpretation. Pragmatics and Argumentation. Cambridge: Cambridge University Press, 2021.

WEINBERGER, D. Our Machines Now Have Knowledge We'll Never Understand. Wired, Boone, 18 abr. 2017. Acesso em: 23 jun. 2021. Disponível em: https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/. Acesso em: 22 nov. 2022.

WEYER, J.; SCHULZ-SCHAEFFER, I. (eds.). Management Komplexer Systeme: Konzepte Für die Bewältigung von Intransparenz, Unsicherheit und Chaos. Berlin: De Gruyter, 2009.

ZUMTHOR, P. Introduction à la poésie orale. Paris: Seuil, 1972.