Inteligência Artificial e o Direito: duas perspectivas

Juliano Maranhão[1]
Juliana Abrusio[2]
Marco Almada[3]

A Inteligência Artificial é cada vez mais relevante para o Direito, graças a duas tendências distintas, mas complementares. A primeira é a crescente adoção de sistemas inteligentes em várias aplicações, tanto na tomada de decisões nos setores públicos e privados quanto na construção de sistemas voltados ao consumidor e assistentes pessoais para as mais diversas tarefas cotidianas. Tal proliferação de inteligências artificiais significa seu envolvimento ubíquo em diversas relações sociais e econômicas tuteladas pelo Direito. Nesse cenário, podemos falar de um Direito da Inteligência Artificial, ou seja, da disciplina jurídica dos agentes digitais e das implicações de seu envolvimento em relações jurídicas e conflitos delas decorrentes.

A segunda decorre do fato de a Inteligência Artificial ser não apenas um objeto externo sujeito à disciplina jurídica, mas também uma ferramenta cada vez mais utilizada por operadores do Direito. Apesar de o emprego de aprendizado de máquina ter apresentado resultados extremamente úteis para advogados e para tribunais, especialmente com análise preditiva, arguiremos aqui que as correlações empíricas nas quais se baseiam enfrentam limitações em sua capacidade explicativa, o que compromete apli-

1 Professor de Direito da Universidade de São Paulo. julianomaranhao@gmail.com
2 Doutora em Direito pela Pontifícia Universidade Católica de São Paulo e professora da Universidade Presbiteriana Mackenzie. juliana@opiceblum.com.br
3 Mestre em Engenharia Elétrica pela Universidade Estadual de Campinas e estudante de Direito na Universidade de São Paulo. marco.almada@usp.br

cações no domínio jurídico, no qual há exigência de justificação normativa das decisões. Acreditamos que a próxima geração de Inteligência Artificial aplicada ao Direito deverá incorporar modelos de representação de conhecimento jurídico às ferramentas que utilizam aprendizado de máquina.

Na próxima seção, apresentaremos os principais desafios envolvidos na regulação da inteligência artificial, cujas consequências jurídicas serão exploradas na terceira seção. Essas duas perspectivas estão relacionadas entre si, já que a disciplina jurídica da Inteligência Artificial influenciará os sistemas construídos para finalidades jurídicas, bem como a tecnologia pode tornar viáveis novas abordagens regulatórias. Portanto, a atuação interdisciplinar entre os profissionais do Direito, da Engenharia e da Computação pode ser benéfica para ambas as áreas, contribuindo para o melhor funcionamento do sistema jurídico e para garantir que as aplicações baseadas em inteligência artificial promovam, ou ao menos não lesionem, os direitos e interesses protegidos pelo Direito brasileiro.

A regulação de sistemas inteligentes

Segundo Cormen et al. (2009), um algoritmo é qualquer procedimento de computador bem definido que possua algum valor agregado na qualidade de suas entradas (*input*), gerando outros valores na saída (*output*), de forma que pode ser considerado uma ferramenta para resolver um problema.

Os algoritmos são empregados em programas de computador por diversas organizações para a tomada de decisões e alocação de recursos, tendo por base grandes conjuntos de dados. Combinando cálculo, processamento e raciocínio, os softwares podem ser excepcionalmente complexos, codificando milhares de variáveis em milhões de pontos de dados. Desse modo, decisões importantes sobre a vida das pessoas são cada vez mais ocasionadas por

sistemas computacionais e algoritmos, como direcionar vagas de empregos e concessão de crédito. A preocupação está em como programas de computador baseados em Inteligência Artificial podem limitar oportunidades e, especialmente, colocar em risco direitos fundamentais dos cidadãos.

Dentre os empregos de inteligências artificiais, um dos principais focos de atenção para o Direito está na coleta e tratamento de dados pessoais para perfilhamento (Caplan et al., 2018[4]). Determinar o perfil do indivíduo pode valer muito a anunciantes, a seguradoras, e ao próprio Estado. A China, por exemplo, país de 1,4 bilhão de habitantes, tem utilizado uma combinação de vigilância por meio de inteligência artificial com uso de enorme quantidade de dados pessoais para monitorar a vida e o comportamento das pessoas em detalhes minuciosos.[5]

---

4 Segundo o artigo 4 (4) da GDPR: "Definição de perfis": qualquer forma de tratamento automatizado de dados pessoais que consista em utilizar esses dados pessoais para avaliar certos aspectos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspectos relacionados com o seu desempenho profissional, a sua situação econômica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações.

5 Para uma visão mais detalhada de como a China tem incorporado uma sociedade de dados para sustentar seus interesses políticos e econômicos, ver Larson (2018). Da reportagem, destacamos: "Nenhum governo tem um plano mais ambicioso e de tanto alcance para utilizar o poder dos dados para mudar a maneira como governa, do que o governo chinês", diz Martin Chorzempa, do Instituto Peterson de Economia Internacional, em Washington, DC [Tradução livre de: *"No government has a more ambitious and far reaching plan to harness the power of data to change the way it governs than the Chinese government," says Martin Chorzempa of the Peterson Institute for International Economics in Washington, DC".]* E ainda sobre o assunto do Sistema de Crédito Social Chinês, por meio de um programa de pontuação sobre os seus cidadãos, classificando-os de acordo com as informações contidas em um enorme banco de dados, alimentado com milhares de informações pessoais (em grande maioria coletadas pela internet), tais como histórico de navegação na internet, itens adquiridos em compras, probabilidade de adimplemento das obrigações etc., ver Gomes (2017, p.50-1). Por fim, importante ressaltar que a tutela à privacidade e proteção de dados na China não é comparável à proteção conferida no Ocidente. Nos últimos anos,

Outra preocupação relevante, da perspectiva jurídica, está na possibilidade de contestar e revisar decisões baseadas em Inteligência Artificial, especialmente quando estejam baseadas *deep learning,*[6] com seu problema de opacidade (black box[7]). A contestabilidade e a possibilidade de revisão pressupõem inteligibilidade, em termos humanos, ou seja, por um conjunto de critérios determinantes que fundamentem determinada decisão.

Não à toa, existem diversas iniciativas no mundo acadêmico jurídico e de ciência da computação para pensar, refletir e propor caminhos sobre o assunto. Dentre outras iniciativas, podem ser ci-

---

entretanto, vários casos foram julgados pelos tribunais chineses sobre o assunto e iniciativas legislativas estão sendo desenvolvidas. Para um panorama da privacidade e proteção de dados na China, ver Ong (2011, p.172-9).

6 *Deep learning* é uma abordagem de aprendizado de máquina que busca resolver problemas a partir da composição entre múltiplos níveis de aprendizado. Essa abordagem obteve bastante êxito nos últimos anos em tarefas como o reconhecimento de fala, a identificação de objetos em imagens e a geração artificial de texto, porém apresenta dificuldades quanto à inteligibilidade de seus resultados, em termos de quais seriam seus critérios determinantes (Goodfellow et al., 2016).

7 *"There is, however, ongoing research into mechanisms for rule extraction, to assist in understanding by extracting knowledge from more opaque approaches and expressing it in a more intelligible form, such as a decision tree. There are also ways to try to describe what aspects of the input led to a particular decision (rather than describing the model as a whole) such as highlighting features of an image that led to a particular classification. These assist in assessing the appropriateness of the model. Control tends to be more challenging for the more opaque models; though there is continuing work on general means for improving and providing control (such as 'fairness') across approaches"* (Singh et al., 2016). Há, no entanto, pesquisas contínuas sobre mecanismos de extração de regras, para auxiliar na compreensão, extraindo conhecimento de abordagens mais opacas e expressando-as de uma forma mais inteligível, como uma árvore de decisão. Há também maneiras de tentar descrever quais aspectos da entrada levaram a uma decisão específica (em vez de descrever o modelo como um todo), como destacar recursos de uma imagem que levaram a uma classificação específica. Esses auxiliam na avaliação da adequação do modelo. O controle tende a ser mais desafiador para os modelos mais opacos; embora haja trabalho contínuo em meios gerais para melhorar e fornecer controle (como "equidade") em várias abordagens.

tadas a Algorithm Watch,[8] da Alemanha; o Lawgorithm,[9] no Brasil, e a International Association for Artificial Intelligence and Law.[10]

A opacidade de sistemas baseados em aprendizado de máquina é uma das maiores fontes de atenção e preocupação na atualidade, especialmente no que diz respeito à possibilidade de

8 Disponível em: <https://algorithmwatch.org/de/>. Acesso em: 24 dez. 2018. O AlgorithmWatch é uma iniciativa dos pesquisadores Lorena Jaume-Palasí, Lorenz Matzat, Matthias Spielkamp e Prof. Dr. med. Katharina Anna Zweig, e é apoiada pelas autoridades da mídia estatal de Hesse, Baviera, Baden-Württemberg, Renânia-Palatinado, Sarre e Saxônia. Tem foco nos algoritmos da gigante Google, bem como nas eleições do país.

9 Lawgorithm é uma associação independente, sem fins lucrativos, fundada por professores de Ciência da Computação, Engenharia, Direito, Economia e Filosofia da USP, dedicada à pesquisa sobre Inteligência Artificial aplicada ao Direito e sobre as implicações jurídicas, econômicas, sociais e culturais da inteligência artificial. Da perspectiva da Inteligência Artificial para o Direito (IA&Direito), o Lawgorithm promove a pesquisa sobre ferramentas computacionais inteligentes capazes de tornar mais eficiente a atuação de juristas e de gerar informações sobre as atividades legislativa e jurisdicional. Da perspectiva do Direito da Inteligência Artificial (Direito da IA), pretende refletir sobre novas questões jurídicas trazidas pela atuação de agentes digitais. O Lawgorithm parte de quatro convicções orientadoras de seus projetos de pesquisa: 1. O raciocínio jurídico é complexo, envolvendo: a) identificação das regras a serem aplicadas; b) o significado dos termos contidos nas regras perante conceitos jurídicos fundamentais c) a adequação das soluções indicadas pelas regras em relação a propósitos de políticas públicas e princípios valorativos; 2. Ferramentas gerais de Inteligência Artificial serão mais eficientes e adequadas quando forem empregadas no Direito com base em representação de conhecimento, análise e inferências típicas dos juristas (conforme convicção 1); 3. Os juristas atuarão com mais qualidade e produtividade quando se desvencilharem de tarefas repetitivas e puderem ter acesso rápido e eficiente ao conhecimento específico necessário ao seu labor intelectual (por ferramentas que satisfaçam 2); 4. A inteligência artificial deve ser compreendida de perspectiva multidisciplinar, considerando suas condições técnicas, impactos econômicos, sociais e culturais, como pressuposto de qualquer regulação ou interpretação de suas implicações jurídicas. (Lawgorithm. Disponível em: <http://www.lawgorithm.com.br/>. Acesso em: 24 dez. 2018).

10 International Association for Artificial Intelligence and Law. Disponível em: <http://www.iaail.org/?q=page/ai-law>. Acesso em: 24 dez. 2018.

contestação, mas também em relação ao risco de incorporação de vieses que resultem em construção de perfis ou tomadas de decisão discriminatórias, ou ainda da possibilidade de tomadas de decisão que ignorem valores humanos ou desrespeitem direitos fundamentais e a dignidade humana.[11]

Frank Pasquale (2015), ao escrever *The black box society:* the secret algorithms that control money and information, aponta que um tipo de situação adversa está em jogo, em que o adversário é a própria regulação em si, instigando a seguinte hipótese: e se os controladores mantiverem, propositadamente, seus feitos de forma opaca, justamente para evitar ou confundir a regulação? Por isso adverte que *"secrecy is approaching critical mass, and we are in the dark about crucial decisions. Greater openness is imperative"* (Pasquale, 2015, p.2).

Para Frank Pasquale (2015) deve haver menor esforço de concentração em tentar controlar a coleta de dados, e mais esforço em regular o uso desses dados – como as empresas e os governos estão realmente implantando regras para tomar decisões, com emprego de inteligências artificiais.

É em relação a esses riscos que se direcionam esforços não só jurídicos, como também de incorporação de critérios éticos no desenvolvimento de sistemas...

A autoridade de proteção de dados da União Europeia, European Data Protection Supervisor, criou, em 3 de dezembro de 2015, o Ethics Advisory Group sobre as dimensões éticas da proteção de dados no atual contexto digital, com processamento de *big data* e uso de inteligência artificial. Em 2018, esse grupo

11 Como mostra Jenna Burrell (2016), a opacidade de um sistema inteligente pode surgir de três fontes principais: a complexidade dos modelos matemáticos envolvidos, a dificuldade de entender as operações envolvidas no processamento de dados em larga escala e a falta de clareza no contexto institucional de uso desses sistemas. A presença de um ou mais desses fatores pode dificultar a identificação de lesões ou ameaças a direitos e interesses, sejam eles individuais ou coletivos, que surjam em decorrência do uso da Inteligência Artificial.

emitiu o Relatório denominado *Towards a Digital Ethics* (Burgess et al., 2018) cujo documento reconhece que a nova ética digital está baseada no direito fundamental à privacidade e à proteção de dados pessoais, entendendo crucial essa observância ética para a proteção da dignidade humana, que é a base da Carta de Direitos Fundamentais da União Europeia. Em abril de 2019, o Parlamento Europeu publicou o estudo intitulado *A governance framework for algorithmic accountability and transparency,*[12] cujo estudo desenvolve e propõe opções de regulações para a governança da transparência algorítmica e de sua responsabilização, com base em análise sob os aspectos sociais, técnicos e regulatórios.

O estabelecimento de princípios éticos para guiar o desenvolvimento da Inteligência Artificial é necessário para conduzir a aplicação da Inteligência Artificial para finalidades socialmente positivas, mas envolve alguns riscos. O primeiro desses é a proliferação de conjuntos de princípios: um estudo recente do Berkman Klein Center for Internet & Society (Fjeld, 2018) mapeou mais de trinta conjuntos de princípios publicados por governos, organizações internacionais, empresas e organizações do terceiro setor, que por vezes adotam princípios distintos uns dos outros ou dão ênfases diferentes aos princípios escolhidos, levando ao problema de identificar quais são os princípios relevantes.

Outra dificuldade está no fato de que os princípios são formulados em discussões *"top down"* e pretensão de universalidade, o que torna seu conteúdo demasiadamente genérico e abstrato, tornando-os de difícil aplicação. Por um lado, a definição de regras gerais em vez de uma regulação rígida é favorável ao desenvolvimento tecnológico e permite que a sociedade se familiarize com os sistemas inteligentes antes de decidir como eles devem ser regulados. Por outro, ela pode contribuir para o que a literatura sobre regulação de inteligência artificial chama de *ethics-washing*

12 Disponível em: <http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf>.

(Wagner, 2018): o uso de princípios vagos como um substituto para regras que efetivamente protegeriam os direitos e interesses individuais e coletivos que poderiam ser afetados pelo uso de sistemas inteligentes.

Assim, os princípios diretores do uso da Inteligência Artificial devem encontrar um equilíbrio entre o excesso e a ausência de regulação. Entendemos que o melhor caminho para discussões éticas que pretendam aplicabilidade seja por meio da análise *"bottom up"*, buscando equilíbrio reflexivo entre princípios gerais e casos concretos em setores específicos. Ou seja, trata-se de discutir não os princípios universais da ética computacional ou algorítmica, mas de desenhar princípios específicos para diferentes setores de aplicação: ética algorítmica no campo da medicina, no campo jurídico, no âmbito comercial etc. Por exemplo, na área de aplicações para medicina, pode ser mais relevante a precisão, ao passo que no direito, a explicabilidade é mais importante.

O equilíbrio ideal dependerá não só das capacidades tecnológicas dos sistemas inteligentes, mas também das peculiaridades de cada domínio de aplicação e das exigências da lei. No caso da Inteligência Artificial aplicada ao Direito, por exemplo, o problema da *"black box"* discutido acima aparece com mais força, já que as decisões judiciais estão sujeitas a uma série de requisitos que ultrapassam a mera acurácia preditiva, como a necessidade de que o conteúdo da decisão possa ser justificado racionalmente. Na próxima seção do texto, exploraremos em mais detalhes os usos jurídicos da inteligência artificial e o estado da arte em relação ao problema de tornar explicáveis as decisões destes sistemas.

A aplicação da Inteligência Artificial ao Direito

Sistemas baseados em Inteligência Artificial são usados,

como vimos acima, para automatizar várias tarefas cuja realização exigiria a inteligência humana. Dentre essas tarefas, é particularmente relevante o uso de sistemas inteligentes para tentar prever o comportamento de indivíduos – o que engloba objetivos diversos, como a identificação das preferências de consumo de alguém e a determinação da probabilidade de inadimplência em um financiamento[13] –, bem como o uso de dados para a tomada de decisão sem a participação humana.

Esses modelos, como discutidos, podem ser opacos ao entendimento humano, o que pode ser resultado de três fontes principais: a complexidade dos modelos matemáticos envolvidos, a dificuldade de entender as operações envolvidas no processamento de dados em larga escala e a falta de clareza no contexto institucional de uso desses sistemas. Como as atividades jurídicas – seja na esfera judicial, seja na esfera administrativa, seja em outros métodos de solução de controvérsias – estão em geral conectadas a situações que podem produzir impactos significativos para pessoas físicas e jurídicas, o ordenamento jurídico impõe uma série de restrições que devem ser seguidas para o uso lícito da Inteligência Artificial como ferramenta de apoio ou de automação.

Uma primeira fonte dessas restrições, como já comentado, é o regramento jurídico da privacidade e da proteção de dados pessoais: os princípios éticos mencionados também se aplicam aos usos jurídicos da Inteligência Artificial, bem como as leis neles baseadas. Em particular, a Lei Geral de Proteção de Dados (LGPD, Lei n.13.709/2018), cuja entrada em vigor está prevista para agosto de 2020, tem dispositivos relevantes para aplicações, como a geração automática de contratos ou a busca automatizada por jurisprudência, que estejam ligadas a questões jurídicas.

A essas preocupações gerais, lidas à luz da discussão da seção

13 Para uma descrição das possibilidades neste sentido, ver Gomes (2017, p.50-1).

anterior, somam-se as questões específicas do âmbito jurídico. Um exemplo é a exigência de que as decisões judiciais sejam fundamentadas na análise das questões de fato e de direito presentes no processo.[14] Diante desse requisito, um sistema inteligente que se proponha a automatizar uma decisão judicial – ou, de forma mais realista para a tecnologia atual, fornecer aportes para um juiz humano – deve ser capaz de fornecer a fundamentação exigida por lei. Porém, a opacidade envolvida no uso e construção dos sistemas inteligentes pode dificultar, ou mesmo tornar inviável, a produção desse tipo de fundamentação.

A LGPD – assim como sua contraparte da União Europeia, o Regulamento Geral de Proteção de Dados (RGPD, ou GDPR na sigla inglesa) – inclui uma forma do chamado direito à explicação.[15] Segundo a LGPD, em seu artigo 20, § 1º, o controlador de sistemas que tomam decisões baseadas unicamente no tratamento automatizado de dados pessoais deve fornecer informações a respeito dos critérios e procedimentos utilizados para a decisão automatizada. Essa formulação da lei se aproxima do paradigma de sistemas baseados em conhecimento, que operam com base em representações predefinidas do conhecimento disponível a respeito do problema que pretendem resolver (Russell; Norvig, 2010).

Muitas das aplicações de Inteligência Artificial que hoje assumem papel de destaque no imaginário popular são, contudo, baseadas em outro paradigma: os sistemas de dados. Essa abordagem, nos últimos anos, foi usada para a construção de sistemas de busca por divergências em acórdãos do Supremo Tribunal Federal (Oliveira, 2017), predição de decisões da Corte Europeia de Direitos Humanos (Medvedeva et al., 2019), e a aplicação de testes

14 Ver, por exemplo, a exigência da fundamentação como elemento essencial da sentença pelo Código de Processo Civil, artigo 489, caput.

15 Para um panorama da discussão a respeito dos limites do direito à explicação na regulação da União Europeia, ver Kaminski (2019). Para uma discussão dos limites de viabilidade da explicação em face das possibilidades tecnológicas, ver Almada (2019).

de propósito principal na tributação internacional, dentre outras aplicações (Kuzniacki, 2018).

Sistemas de dados não são construídos com base em representações predefinidas do problema que pretendem responder. Em vez disso, eles operam com base no aprendizado de máquina (Haykin, 2008): antes de serem usados para sua aplicação, esses sistemas são expostos a dados sobre o problema a resolver, e por meio de processos estatísticos de treinamento[16] extraem da base de dados padrões e correlações que serão generalizados para resolver problemas futuros.

Esses métodos de predição são empíricos, ou seja, extraem os resultados de uma série decisões judiciais e as correlacionam com fatores como o tipo de demanda, valor envolvido, e o tribunal em que a demanda é julgada. Como sua estrutura é baseada nas propriedades detectáveis a partir dos dados, os sistemas baseados em aprendizado de máquina não levam em consideração qualquer justificação normativa sobre como deve ser a decisão a partir das características e argumentos do caso.

Além disso, a complexidade matemática dos modelos empregados para extrair as propriedades dos dados torna inviável, mesmo para o observador técnico, o detalhamento minucioso de como estes sistemas operam. Quando essas técnicas são aplicadas em cenários de *big data,*[17] a escala das operações envolvidas torna o cenário ainda mais complexo, dificultando a exposição das decisões

16 O processo de treinamento pode ser supervisionado, em que o sistema aprende a associar determinados valores dos dados de entrada com rótulos fornecidos para seu treinamento; de reforço, quando as respostas do sistema são treinadas com base na recompensa fornecida a ele por uma resposta correta; ou não-supervisionado, em que o treinamento busca alinhar o sistema inteligente com uma métrica que reflete propriedades dos dados e não um resultado desejado (Haykin, 2008, p.34-45.)

17 Parte da popularização deste paradigma da inteligência artificial nos últimos dados está ligado ao chamado big data, isto é, a formação de vastas bases de dados e as subsequentes possibilidades de extração de informação a partir destes dados. Sobre o tema, ver Gomes (2017).

tomadas pelo sistema em forma compreensível para humanos, com premissas, critérios acessíveis, argumentos e conclusões.[18]

Nesta seção, discutiremos os êxitos e limitações de ambos os paradigmas do aprendizado de máquina. De um lado, os sistemas baseados em conhecimento conseguem representar uma vasta gama de aplicações jurídicas, bem como produzir respostas inteligíveis, mas sua construção e uso exige um esforço que muitas vezes dificulta aplicações práticas. Já os sistemas de aprendizado de máquina conseguem extrair correlações estatísticas com menos estrutura do que os sistemas baseados em conhecimento, mas suas soluções não são explicáveis em um nível compatível com as exigências jurídicas nem se beneficiam do conhecimento jurídico. Por fim, defenderemos a combinação entre técnicas baseadas em conhecimento e técnicas de aprendizado de máquina, aproveitando os pontos positivos de ambas as abordagens para a construção de sistemas mais adequados às peculiaridades das predições jurídicas, de natureza normativa.

Sistemas baseados em conhecimento jurídico

Sistemas baseados em conhecimento são sistemas inteligentes que realizam inferências com base em representações internas de conhecimento (Russell; Norvig, 2010). Essas representações do conhecimento são construídas de antemão, durante o projeto de um sistema computacional, mas uma vez que elas sejam feitas, o sistema pode atualizar sua base de conhecimentos a partir das informações que adquire do ambiente, aplicando as mesmas regras de inferência para lidar com situações percebidas.

No caso dos sistemas baseados em conhecimento jurídico, esse raciocínio opera através da estrutura formal, que pode ser construída a partir de diferentes lógicas, como lógicas deônticas

18 Sobre ambas as questões deste parágrafo, ver Burrell (2016).

(Hilpinen; McNamara, 2013), que lidam com conceitos como obrigações, proibições e permissões, seja por operadores modais, seja por meio de representações de conjuntos de normas, constitutivas, obrigatórias ou permissivas, como nas lógicas *input/output*, ou ainda, e lógicas de argumentação derrotável.

Lógicas de argumentação derrotável encontraram ampla aplicação em representação de conhecimento jurídico, seja o raciocínio baseado em precedentes (*case based reasoning*), seja o raciocínio baseado em regras (*rule based reasoning*). Essas lógicas modelam o raciocínio como inferências a partir de argumentos a favor ou contra determinada tese, incluindo informação sobre a força relativa destes argumentos. Um argumento pode ser uma estrutura inferencial complexa, que liga suas premissas a conclusões por meio de passos intermediários detalhados.[19] Esses argumentos, por sua vez, podem ser atacados em diferentes junções e de diversas formas, e uma conclusão pode ser derivada se for possível construir um argumento a favor da conclusão que seja defensável contra todos os argumentos que o atacam.

Tal estrutura baseada em argumentos, contra-argumentos, refutações e presunções alinha-se de forma direta com o raciocínio jurídico, tendo em vista que decisões judiciais são tomadas a partir da avaliação de argumentos das partes em oposição. Com isso, as lógicas de argumentação foram aplicadas com sucesso para representar vários aspectos do raciocínio jurídico. Dentre as aplicações bem-sucedidas das lógicas de argumentação jurídica na Inteligência Artificial, temos modelos de raciocínio sobre argumentação oral em cortes (Bench-Capon; Prakken, 2010), evidências processuais (Bex et al., 2003; Verheij, 2003) e raciocínio com precedentes judiciais. Em particular, mais recentemente, modelos lógicos de argumentação foram desenvolvidos nos quais

19 De um ponto de vista técnico, tanto os passos intermediários quanto a conclusão final devem ser autorizados por operações de uma lógica dedutiva ou derrotável (não monotônica).

ocorre o balanço de múltiplas considerações prós e contras (fatores do caso, razões, princípios e valores). Além disso, alguns modelos formulam princípios racionais sobre o desenvolvimento dos precedentes no tempo e sobre a dinâmica de construção e alteração de conceitos em interpretações de leis e precedentes judiciais (Hage, 2005; Sartor, 2013; Prakken et al., 2013; Horty, 2011).

Boa parte dos avanços de IA & Direito sobre a interpretação de conceitos gira em torno da noção de "fator", que surgiu em dois programas de computador pioneiros em IA & Direito: o sistema HYPO[20] e o sistema CATO.[21] Fatores são abstrações ou estereótipos da descrição de um caso, que podem favorecer (fatores favoráveis) ou prejudicar (fatores contrários) uma conclusão legal. Por exemplo, o fator "gravidez resultante de estupro" é um fator favorável, no direito brasileiro, à decisão de permitir ou não punir o aborto. Já o fator "procedimento não realizado por médico" é um fator que leva à proibição do aborto.

A atenção dada ao papel dos valores e propósitos levou a abordagens sobre a interpretação jurídica como um problema de decisão, ou seja, como uma escolha entre interpretações alternativas considerando a probabilidade daquela interpretação ou de decisões baseadas naquela interpretação, ou ainda, de consequências possíveis daquela interpretação em termos de promoção ou demoção dos valores relevantes. Naquelas abordagens, a escolha de uma interpretação é baseada nos efeitos positivos e negativos que as potenciais decisões podem ter em relação a valores ou objetivos que sejam tomados como relevantes, considerando as preferências relativas entre esses valores ou objetivos, e, em alguns modelos, também considerando a extensão na qual esses valores ou objetivos são alcançados ou impactados. Há também ligações entre essas abordagens e teorias da decisão qualitativas (Keeney;

20 Introduzido por Ashley (1990).

21 Introduzido por Aleven (2003, p.183ss.).

Raiffa, 1976), recentemente exploradas por Giovanni Sartor.[22]

Sistemas computacionais mais recentes, como o VJAP (Grabmair, 2017), desenvolvido para aplicação em temas de concorrência desleal, em particular sobre violação de segredo de indústria (*trade secret law*), procuram incorporar valores e propósitos perseguidos na construção de justificações para decisões judiciais. Com isso, o sistema faz predição do resultado do caso, ou seja, da provável decisão judicial, por meio de uma medida de confiança derivada dos grafos argumentativos e gera textos com argumentos justificando a predição.

Em todas essas abordagens, é possível reconstruir a lógica que o sistema utilizou para construir suas predições, o que fornece uma justificação racional para as decisões tomadas. Contudo, não só a construção destes sistemas exige o emprego de muito conhecimento específico sobre o tema, como também há a necessidade de se extrair manualmente a informação que os algoritmos usarão, o que exige, na prática, um trabalho substancial de pré-processamento para a construção e uso desses sistemas. Por isso, os sistemas baseados em conhecimento não são, hoje, tão aplicados quanto a sua sofisticação teórica poderia sugerir.

Aplicações jurídicas do aprendizado de máquina

Nos últimos anos, técnicas de aprendizado de máquina foram empregadas para prever resultados de decisões judiciais a partir de textos jurídicos. Um exemplo é o modelo construído por Nikolaos Aletras et al. (2016), que atingiu 79% de acurácia ao avaliar se a Corte Europeia de Direitos Humanos decidiria pela existência de uma violação de direitos em um determinado técnico. Em aplicações jurídicas, abordagens baseadas em Máquinas de Veto-

22 Dentre a produção deste autor sobre o tema, destacamos dois artigos: "Fundamental legal concepts: A formal and teleological characterisation" e "The logic of proportionality: reasoning with non-numerical magnitudes" (Sartor, 2010; 2013).

res-Suporte (SVM) têm obtido os melhores resultados, embora abordagens baseadas em *deep learning* venham ganhando espaço (Contissa et al., 2018).

De forma geral, esses modelos baseiam suas predições em elementos textuais (*features*) que revelam padrões para as decisões. A construção de um modelo de aprendizado de máquina para o processamento de linguagem natural – o que inclui o processamento de textos jurídicos – envolve, em geral, quatro passos: (1) a compilação de um *corpus* de textos relevantes para o domínio da aplicação; (2) o pré-processamento dos textos desse *corpus*, para deixá-los em um formato que os algoritmos de processamento de linguagens naturais possam consumir; (3) a anotação desses textos, por meios automáticos ou manuais, para atribuir rótulos adequados (por exemplo, para dizer se um recurso foi provido); e (4) o treinamento do modelo que realizará as previsões desejadas (Eckhardt de Castilho, 2018). Todas essas etapas podem se beneficiar do uso de conhecimentos específicos do domínio de aplicação como uma forma de melhorar o tratamento computacional dos textos analisados, como fazem sistemas como CLAUDETTE, desenvolvido na Universidade de Bologna, que detecta cláusulas abusivas em documentos que descrevem ou pedem concordância com as políticas de privacidade de sites (Contissa et al., 2018).

Mesmo com o auxílio de conhecimento especialista, todavia, os padrões encontrados por um sistema inteligente muitas vezes são aqueles que um humano observaria ao desempenhar a mesma tarefa. Isso porque a predição normativa, típica do jurista, não se baseia na correlação empírica entre eventos, mas na apreciação dos fatores (características) do caso à luz das normas jurídicas aplicáveis, ou dos argumentos a favor ou contra determinada pretensão. Portanto, difícil é difícil para o jurista humano entender quais são os eventos empíricos que o sistema julgou como relevantes e as correlações utilizadas que resultam no output do sistema.

Exemplo nesse sentido pode ser encontrado no trabalho de Verma et al. (2017), que desenvolveu um sistema com cerca de 75% de acurácia para predizer quando juízes das cortes de apelação norte-americanas divergiriam. A hipótese seria que a divergência seria determinada por diferenças de natureza ideológica, porém, os fatores que mais correlacionaram na determinação do resultado foram: (i) a posição em que os juízes se sentavam no julgamento; (ii) o tamanho dos votos e (iii) o número de citações de precedentes. Obviamente, na análise jurídica, fatores como esses são absolutamente irrelevantes. A predição jurídica de divergência liga-se à avaliação de convicções sobre teses jurídicas e princípios aceitos pelos juízes sob análise.

Assim, a pesquisa em inteligência artificial tem buscado o desenvolvimento de sistemas de inteligência artificial inteligíveis (Explainable Artificial Intelligence: xAI), tanto por meio da produção de sistemas capazes de explicar de forma mais simples o funcionamento de outros sistemas quanto pela construção de sistemas capazes de atingir bom desempenho a partir de mecanismos internos que incorporem representação de conhecimento jurídico.

Abordagens híbridas e o problema da explicação das decisões

Apesar de seus êxitos, ambos os paradigmas de Inteligência Artificial encontram limites à sua aplicação em questões jurídicas. No caso dos sistemas baseados em conhecimento, o principal obstáculo consiste na dificuldade envolvida na codificação dos dados de um caso para que eles tenham uma forma que o sistema possa entender e processar, trabalho que hoje é majoritariamente feito por humanos. Para sanar essa dificuldade, há linhas prósperas de pesquisa voltadas ao desenvolvimento de sistemas para identificação automática de fatores juridicamente relevantes em textos jurídicos por meio de emprego de espaços vetoriais (Ashley;

Falakmasir, 2017) e de detecção automática e estruturação de argumentos.[23] Embora essas tecnologias ainda precisem atingir maior grau de maturação, elas já contribuem para reduzir o trabalho humano necessário para a adoção de sistemas baseados em conhecimento jurídico.

Da mesma forma que o uso de técnicas de aprendizado de máquina pode servir para viabilizar o uso prático de sistemas baseados em conhecimento, as representações do domínio que estes proporcionam podem ser úteis para explicar como a inteligência artificial jurídica chega às suas conclusões. No campo do Direito, a IA inteligível é de particular importância, uma vez que qualquer ato ou decisão judicial ou administrativa somente é juridicamente válido na medida em que possa ser juridicamente justificada. Nesse domínio, o processo e conteúdo de justificação é tão relevante quanto o resultado.

Como exemplos de projetos que buscam fornecer explicações para decisões automatizadas no âmbito jurídico, temos o projeto NWO Forensic Science,[24] coordenado por Bart Verheij, que busca desenvolver sistemas para gerar explicações sobre redes bayesianas e modelos probabilísticos de análise de evidências processuais, por meio de cenários e argumentos de forma que os modelos se tornem compreensíveis para juristas.[25]

A necessidade de explicação não é apenas um requisito para que os sistemas inteligentes possam ser utilizados em aplicações jurídicas, mas é também um requisito para que as predições da inteligência artificial sejam eficazes. Posto que uma decisão judicial muitas vezes não é uma consequência imediata do "estímulo dos fatos" (Aletras, 2016) que possam ser detectados diretamente a partir do texto, uma predição de decisões judiciais não deve se limitar aos aspectos empíricos, mas também levar em conta as

23 Nesse sentido, ver Palau e Moens (2009), bem como Pathak, Goyal e Bhowmick (2016).

24 Disponível em : <http://www.ai.rug.nl/~verheij/nwofs/>.

25 Nesse sentido, ver Verheij et al. (2016), bem como Vlek (2014).

dinâmicas argumentativas que devem ser reconstruídas para que um sistema possa fazer uma predição adequada. Dessa forma, a incorporação de técnicas computacionais de tratamento da argumentação jurídica às técnicas já existentes de processamento de linguagem jurídica permitirá que os sistemas consigam identificar os argumentos presentes em um documento jurídico e, a partir disso, prever o resultado de uma forma ao mesmo tempo mais precisa e passível de explicação para os observadores humanos.

O caminho para a integração de modelos de predição é, de um lado, o desenvolvimento de programas capazes de identificar fatores relevantes ou argumentos em documentos jurídicos de tal modo que possam gerar premissas para sistemas computacionais baseado em lógicas jurídicas e de argumentação; e, de outro, o desenvolvimento e refinamento dos sistemas de argumentação e construção de justificações com base em fatores e argumentos de modo a que possam processar argumentos e fatores extraídos automaticamente (por aprendizado de máquina) de textos e documentos jurídicos. Com isso, é possível desenvolver modelos de argumentação que possam, a partir de textos jurídicos com pouco ou nenhum tratamento especial, oferecer não só predições empíricas, baseadas em correlações estatísticas, como também predições normativas, extraídas da força argumentativa do caso em relação a precedentes, bem como a construção de justificações jurídicas a partir de argumentos inteligíveis.

Considerações finais

A difusão do uso de sistemas inteligentes tem o potencial de transformar a prática do Direito, não só por trazer novas questões a serem consideradas as profissões jurídicas, mas também pela automação de atividades jurídicas, começando por aquelas que envolvem trabalho repetitivo e posteriormente se sofisticando. Essas mudanças exigirão uma mudança no perfil do profissional

jurídico, que precisará estar apto a lidar com o novo cenário social e com as novas tecnologias.

No que tange as aplicações da inteligência artificial ao direito, esse processo de adaptação envolve tanto o uso das ferramentas de inteligência artificial pelos profissionais do Direito quanto o aproveitamento de seu conhecimento para a construção de sistemas computacionais que sejam capazes de realizar suas atividades de tratamento de dados de formas compatíveis com a lei. Portanto, a popularização da Inteligência Artificial exigirá profissionais capazes de lidar com as transformações tecnológicas e de operar em equipes interdisciplinares, que aproveitem as competências de juristas, cientistas da computação e outros profissionais para a construção de sistemas inteligentes que tenham efeitos positivos e protejam os direitos e interesses juridicamente tutelados em jogo.

Referências

ALETRAS, N. et al. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science*, v.2, 2016.

ALEVEN, V. Using background knowledge in case-based legal reasoning: a computational model and an intelligent *learning* environment. *Artificial Intelligence*, n.150, 2003.

ALMADA, M. Human intervention in automated decision-making: Toward the construction of contestable systems. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (Icail 2019). Montreal, Canadá, 2019, seção 3.

ASHLEY, K. D. *Modeling Legal Argument*: Reasoning with Cases and Hypotheticals. Cambridge: MIT Press, 1990.

ASHLEY, K.; FALAKMASIR, M. H. Utilizing Vector Space Models for identifying legal factors from text. *Proceedings of Jurix Conference*, 2017.

BENCH-CAPON, T. J. M.; PRAKKEN, H. Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, v.18, 2010.

BEX, F. J. et al. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, v.11, p.125-65, 2003.

BURGESS, P. et al. Ethics advisory group. Towards a digital ethics. 2018 Disponível em: <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf>.

BURRELL, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, v.3, n.1, p.1-12, 2016.

CAPLAN, R. et al. Algorithmic accountability: a primer. *Data & Society,* 18 abr. 2018. Disponível em: <https://datasociety.net/output/algorithmic-accountability-a-primer/>.

CONTISSA, G. et al. CLAUDETTE meets GDPR: automating the evaluation of privacy policies using Artificial Intelligence. *Relatório Técnico para BEUC*, 2018.

CORMEN, T. et al. *Introduction to algorithms*. 3.ed. Massachusetts: The MIT Press, 2009.

ECKHART DE CASTILHO, R. A Legal Perspective on Training Models for Natural Language Processing. In: LREC 2018, Miyazaki, Japão, 2018.

FJELD, J. et al. Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches. *Relatório técnico, Berkman Klein Center For Internet & Society,* 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, *A. Deep learning.* Cambridge: MIT Press, 2016.

GOMES, R. D. P. *Big Data*: desafios à tutela da pessoa humana na sociedade da informação. Rio de Janeiro: Lumen Juris, 2017.

GRABMAIR, M. Predicting trade secret case *outcome*s using argument schemes and learned quantitative value effect tradeoffs. In: *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law* (Icail). [s.l.], 2017.

HAGE, J. C. Comparing alternatives in the law. *Artificial Intelligence and Law*, v.12, 2005.

HAYKIN, S. *Neural Networks and Learning Machines*. [s.l.]: Prentice Hall, 2008.

HILPINEN, R.; MCNAMARA, P. Deontic Logic: a historical survey and introduction. In: GABBAY, D. et al. *Handbook of Deontic Logic and Normative Systems.* [s.l.]: College Publications, 2013.

HORTY, J. F. Rules and reasons in the theory of precedente. *Legal Theory*, v.17, 2011.

KAMINSKI, M. The Right to Explanation, Explained. *Berkeley Technology Law Journal*, v.34, n.1, 2019.

KEENEY, R. L.; RAIFFA, H. *Decisions with Multiple Objectives*. New York: Wiley, 1976.

KUZNIACKI, B. The Artificial Intelligence Tax Treaty Assistant: Decoding the Principal Purpose Test. *Bulletin for International Taxation*, v.72, n.9, 2018.

LARSON, C. Who needs democracy when you have data? T*echnology Review*, 2018. Disponível em: <https://www.technologyreview.com/s/611815/who-needs-democracy-when-you-have-data/>. Acesso em: 27 ago. 2018.

MEDVEDEVA, M.; VOLS, M.; WIELING, M. Using *machine learning* to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 2019.

OLIVEIRA, R. B. de. *Utilização de Ontologias para Busca em Base de Dados de Acórdãos do STF*. São Paulo, 2017. Dissertação

(Mestrado em Ciências da Computação) – Instituto de Matemática e Estatística, Universidade de São Paulo.

ONG, R. Recognition of the right to privacy on the Internet in China. *International Data Privacy Law*, v.1, 3.ed. ago. 2011, p.172–179. Disponível em: <https://doi.org/10.1093/idpl/ipr008>. Acesso em: 10 out. 2018.

PALAU, R. M.; MOENS, M.-F. Argumentation Mining: the detection, classification and structuring of Arguments in text. In: *Proceedings of Icail 2009,* Barcelona, 2009.

PASQUALE, F. *The black box society*: the secret algorithms that control money and informtion. Cambridge, Ma.: Harvard University Press, 2015. p.2.

PATHAK, A.; GOYAL, P.; BHOWMICK, P. A two-phase approach towards identifying argument structure in Natural Language. In: *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Aplications*. Osaka, 2016, p.11-19.

PRAKKEN, H. et al. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 2013.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence*: a modern approach. 3.ed. Upper Saddle River: Prentice Hall, 2010.

SARTOR, G. Fundamental legal concepts: A formal and teleological characterisation. *Artificial Intelligence and Law*, v.21, p.101-42, 2010.

_____. The logic of proportionality: reasoning with non-numerical magnitudes. *German Law Journal*, v.14, 2013.

SINGH, J. et al. Responsibility and *machine learning*: part of a process. 27 out. 2016. Disponível em: <https://ssrn.com/abstract=2860048>.

VERHEIJ, B. Dialectical argumentation with argumentation schemes: an approach to legal logic. *Artificial Intelligence and Law*, v.11, p.167-95, 2003.

VERHEIJ, B. et al. Arguments, Scenarios and Probabilities: Connections Between Three Normative Frameworks for Evidential Reasoning. *Law, Probability & Risk*, v.15, p.35-70, 2016.

VERMA, S.; PARTHASARATHY, A.; CHEN, D. The Genealogy of ideology: predicting agreements and Persuasive memes in the U.S. Courts of Appeals (Icail). London: ACM Press, 2017.

VLEK, C. S. Building Bayesian Networks for Legal Evidence with Narratives: a Case Study Evaluation. *Artificial Intelligence and Law,* v.22, n.4, p.375-421, 2014.

WAGNER, B. Ethics as an Escape from Regulation:From ethics--washing to ethics-shopping? In: HILDEBRANDT, M. (Ed.) *Being Profiling*. Cogitas ergo sum. Amsterdam: Amsterdam University Press, 2018.