

# DECISÕES JUDICIAIS COMPUTACIONALMENTE FUNDAMENTADAS: uma abordagem a partir do conceito de *explainable artificial intelligence*

COMPUTATIONAL LEGAL REASONING: an approach based in explainable artificial intelligence concept

Gustavo Xavier de Camargo<sup>1</sup>

---

Artigo aceito como convidado.

## Resumo

Este artigo aborda a possibilidade e validade da utilização de aplicações desenvolvidas a partir de tecnologias de aprendizagem de máquina (*machine learning*) para a elaboração de decisões judiciais computacionalmente fundamentadas, a partir da equiparação das decisões humanas e computacionais quanto à dificuldade de estabelecimento de um método verificável capaz de garantir, a priori, integridade e correção, sendo a efetiva fundamentação a única forma de controle válida para ambos os tipos de decisão.

## Palavras-chave

decisão jurídica; inteligência artificial; hermenêutica computacional; computação cognitiva; pragmatismo jurídico; fundamentação das decisões.

## Abstract

This article addresses the possibility and validity of the use of machine learning applications for the writing of computer-based court decisions, comparing human and computational decisions regarding the difficulty of establishing a verifiable method that guarantees, a priori, integrity and correctness, and concluding that effective legal reasoning is the only valid form of control for both types of decision.

## Keywords

spelled judgement; artificial intelligence; computational hermeneutics; cognitive computing; legal pragatism; legal reasoning.

---

<sup>1</sup> Mestrando em Direito pela Universidade Federal de Santa Catarina (UFSC). Bacharel em Ciência da Computação pela Universidade Estadual de Campinas (UNICAMP). Bacharel em Direito pela UNOCHAPECÓ. MBA Executivo pelo Insper. Pós-graduado (lato sensu) em Direito Digital e Compliance pela Faculdade Damásio. E-mail: gustavo@xavierdecamargo.com.br.

## 1 Introdução

Em agosto de 2015, o Desembargador Cesar Cury, presidente do Núcleo Permanente de Métodos Consensuais de Solução de Conflitos (Nupemec) do Tribunal de Justiça do Estado do Rio de Janeiro, anunciou, em artigo publicado na imprensa, a criação das Câmaras de Solução On-line (ODRs), tendo como foco a resolução pré-processual dos conflitos, principalmente aqueles relativos a relações de consumo envolvendo concessionárias de serviços públicos, empresas do setor financeiro e planos de saúde, a partir de uma plataforma digital (CURY, 2015).

Já em março de 2018, as Câmaras de Solução On-line voltaram à mídia, agora como alvos de protesto da Ordem dos Advogados do Brasil, do Rio de Janeiro, em decorrência da informação de que a aplicação utilizada para suportar as ODRs utilizaria “robôs” e inteligência artificial para subsidiar “o pedido de um reclamante com mais informações” (PÁDUA, 2018) a partir do histórico de decisões do tribunal sobre determinado caso. Segundo a OAB-RJ, em nota oficial publicada em seu site, tal medida “exclui a participação de advogados na resolução de conflitos, já que os casos serão analisados via internet por um robô. O programa de informática será financiado pelas empresas mais demandadas” (OAB/RJ, 2018).

A edição especial Exame CEO da Revista Exame, publicada em maio de 2018, integralmente dedicada à inteligência artificial, traz matéria assinada pelo Dr. Juliano Maranhão, professor na Faculdade de Direito da USP, sobre os impactos desta tecnologia no sistema de justiça (MARANHÃO, 2018). Ao tratar do tema, destaca a aprendizagem de máquina (machine learning) como a área da inteligência artificial com maior potencial de impacto no setor jurídico e que tais tecnologias podem ser utilizadas tanto em escritórios de advocacia quanto pelo próprio Poder Judiciário. Na sua visão:

enquanto os juristas aliam conhecimento com reflexão para avaliar uma quantidade restrita de documentos, os agentes artificiais são capazes de analisar velozmente milhares de documentos, encontrando padrões que os humanos dificilmente conseguem enxergar. Assim, a inteligência artificial tende não somente a realizar tarefas rotineiras de modo mais eficiente como também a criar novos tipos de serviço e até mesmo novas formas de compreender o direito.

Na mesma esteira, a Receita Federal do Brasil anunciou que utilizará inteligência artificial no julgamento de processos administrativos fiscais de pequeno valor (até R\$ 20.000,00), de modo a dar vazão ao imenso acervo de processos acumulados. Também em matéria em veículo de grande circulação, de abril de 2018, “André Rocha Nardelli, coordenador-geral de Contencioso Administrativo e Judicial da Receita Federal, diz que essa é a primeira vez que a Receita usa a chamada “computação cognitiva” para auxiliar no julgamento de processos” (PUPO; OLIVON, 2018). A matéria destaca, também, que a aplicação deverá ler os autos, identificar as alegações da defesa e chegará a elaborar propostas de decisão.

Este pequeno levantamento de matérias veiculadas na imprensa, entre os meses de março e maio de 2018, tem como objetivo destacar não apenas a importância que vem tomando a preocupação acerca do uso de inteligência artificial no âmbito do processo, seja judicial ou administrativo, mas principalmente o imenso apelo do tema junto à opinião pública. Não se trata, obviamente, de questão reclusa aos debates jurídicos e acadêmicos, mas sim de assunto de interesse geral que, ao fim e ao cabo, refere-se à aceitação, pela sociedade, do fato de termos nossas ações julgadas não por outro ser humano, mas por uma máquina.

A questão, porém, mais importante acerca deste problema não está em definir se sofreremos o impacto do julgamento das máquinas, pois isso já acontece hoje, em inúmeras situações da vida cotidiana. Um exemplo claro destes julgamentos automatizados, com evidente potencial de restrição de direitos, é a utilização de sistemas de *credit scoring* para decisão de concessão de crédito, que cria verdadeiro sistema de “classificação” de consumidores a partir de análise de dados pessoais, de forma a criar uma avaliação probabilística de cumprimento das obrigações, por parte do consumidor, decorrentes da concessão do crédito. A questão realmente relevante acerca das decisões computacionais em processos administrativos ou judiciais relaciona-se a como seremos capazes de avaliar a qualidade da decisão gerada por computador, o que se dará, inexoravelmente, pela avaliação da qualidade de fundamentação destas decisões.

Os diversos trabalhos acerca da forma de decisão dos juízes, que levam em consideração os fatores extrajurídicos que influenciam as decisões judiciais, mostram a imensa complexidade em se estabelecer um modelo capaz de mapear a forma como, verdadeiramente, as decisões são concebidas pelos juízes humanos. Portanto, as questões relativas à obscuridade do modo de tomada de decisão por sistemas computacionais, principalmente aqueles baseados em computação cognitiva parecem não destoarem da obscuridade que cerca a forma de concepção das decisões judiciais humanas.

O trabalho ora apresentado, inicialmente explora esta similaridade entre os processos de decisão humano e computacional, identificando, para ambos os casos, a impossibilidade de desenvolvimento de um método capaz de garantir integridade e correção a partir da análise do procedimento de concepção da decisão. Em seguida, apresenta de forma sucinta os principais conceitos relacionados a inteligência artificial, com destaque para o que se convencionou denominar de computação cognitiva. Ao final, apresenta o conceito de *Explainable Artificial Intelligence* (XAI) como alternativa a produção de decisões jurídicas computacionalmente fundamentadas.

## 2 Pragmatismo do programador e pragmatismo jurídico

Um dos maiores cientistas da computação de todos os tempos, o americano Ken Thompson foi, juntamente com Dennis Ritchie, o criador do sistema operacional Unix. Também é criação de Ken, como é conhecido no meio técnico, a linguagem de computação B, da qual se originou a amplamente utilizada linguagem de programação C, da qual também é um dos criadores. Pelas suas grandes contribuições à ciência da computação, Ken Thomson recebeu, em 1983, o Prêmio Turing, a maior honraria do mundo na área da ciência da computação, algo equivalente ao Prêmio Nobel dessa área.

No discurso de premiação, intitulado, numa tradução livre, *Reflexões sobre a Confiabilidade da Confiança*<sup>2</sup> (THOMPSON, 1984), Ken apresentou um singelo trecho de código, geralmente utilizado como exercício de programação, cuja função é se “re-escrever” sempre que é executado, ou seja, a aplicação reescreve o próprio código que o gerou como resultado de sua execução. Em seguida, discorreu sobre a possibilidade de se inserir este trecho de código dentro de um compilador, um tipo de programa que traduz outros programas para a linguagem interna do computador, ou seja, um software que gera outros softwares. No passo seguinte, apresenta a ideia de que o código que se reescreve poderia ser um determinado tipo de vulnerabilidade, um cavalo de Tróia, que permitiria

<sup>2</sup> No original: “Reflections on trusting trust” (tradução livre)

que uma pessoa fizesse o login em um sistema Unix sem possuir a senha. Discorre, em seguida, com maiores detalhes sobre a forma como esta combinação de códigos poderia “esconder” a vulnerabilidade e conclui que, uma vez gerado um compilador com o código malicioso, é possível apagar os rastros da “infecção” eliminando o trecho de código que a implantou no compilador. Assim, o compilador gerará programas contaminados sem que os programadores percebam, mesmo que novas versões do compilador sejam desenvolvidas a partir do código “descontaminado”.

Ao final, apresenta a seguinte moral da história (THOMPSON, 1984):

A moral é óbvia. Você não pode confiar em um código que não foi totalmente criado por você. (Especialmente código de companhias que empregam pessoas como eu). Por mais que você faça testes ou examine cuidadosamente o código, isso não irá impedir que você use programas contaminados.<sup>3</sup>

Na mesma linha, pode-se abordar as decisões judiciais. É possível garantir que uma decisão judicial seja estabelecida a partir de uma atribuição de sentido decorrente de um esforço hermenêutico autêntico, onde se obtenha, como resultado final, a decisão? Em outras palavras, é possível que o processo de decisão, realizado por um ser humano, não esconda (da mesma forma como o programador “esconde” seu código malicioso) uma intenção previamente definida, fruto do solipsismo daquele que decide, sob uma fundamentação construída a posteriori?

Richard Posner, em seu livro clássico “Como os juízes pensam”, responde negativamente a estas questões, ao abordar o caráter inescapável das decisões pragmáticas, para defender sua posição francamente favorável ao pragmatismo (POSNER, 2008):

O cerne do pragmatismo jurídico é a decisão pragmática, e o cerne desta está no aumento da preocupação judicial com as consequências e, portanto, a disposição política de basear suas decisões nessas consequências. Mas, ao invés de ser sinônimo de decisão ad hoc, no sentido de ter em conta apenas as consequências para as partes no caso concreto, o pragmatismo jurídico sensato diz ao juiz para considerar as consequências sistêmicas, inclusive institucionais, além das consequências diretas do caso concreto. Assim, o juiz deve considerar os efeitos sobre a atividade comercial ao desconsiderar a redação literal de um contrato ou de não aderir a precedentes legais nos quais a comunidade comercial passou a confiar.<sup>4</sup>

Neste sentido, muito didática a lição de Patrícia Perrone acerca da decisão pragmática dos juízes, inclusive ao escolher uma abordagem legalista ou fundada nas consequências na elaboração da decisão de determinado caso (MELLO, 2015):

o magistrado poderá adotar como postura a observância de normas e precedentes, por entender que o custo-benefício de adoção do comportamento legalista, como regra, é superior à adoção de uma ampla liberdade decisória. O comportamento legalista tem o efeito de garantir a previsibilidade do direito e o tratamento isonômico entre os jurisdicionados e poupa trabalho aos magistrados que, ao se apoiarem em precedentes para o julgamento de novos casos, reduzem seu ônus argumentativo. No entanto, diante

<sup>3</sup> No original: “The moral is obvious. You can't trust code that you did not totally create yourself. (Especially code from companies that employ people like me.) No amount of source-level verification or scrutiny will protect you from using untrusted code.” (tradução livre)

<sup>4</sup> No original: “The core of legal pragmatism is pragmatic adjudication, and its core is heightened judicial concern for consequences and thus a disposition to base policy judgments on them rather than on conceptualisms and generalities. But rather than being a synonym for ad hoc adjudication, in the sense of having regard only for the consequences to the parties to the immediate case, sensible legal pragmatism tells the judge to consider systemic, including institutional, consequences as well as consequences of the decision in the case at hand. He thus must consider the effects on commercial activity of disregarding the actual wording of a contract or failing to adhere to legal precedents on which the commercial community has come to rely.” (tradução livre)

de casos difíceis e inéditos ou de decisões que poderiam gerar consequências absurdas, as fontes legalistas são, para o juiz pragmático, apenas um dos múltiplos aspectos a serem levados em consideração para um julgamento e não têm primazia sobre os demais. O elemento predominante em tal reflexão são as consequências de adotar esse ou aquele entendimento. A valoração de tais consequências, por sua vez, é produto de uma complicada interação de modos de raciocinar, da intuição, da emoção, do senso comum, das inclinações políticas e ideológicas, dos traços de personalidade, das experiências pessoais e profissionais de cada juiz. É produto, portanto, de uma decisão que leva em conta todo e qualquer argumento que seja relevante para se definir o melhor resultado a ser alcançado no caso específico – uma abordagem que poderia ser denominada, segundo Tushnet, “*all things considered*”.

Estas abordagens demonstram que, mesmo no campo das decisões judiciais empreendidas por seres humanos, uma abordagem eminentemente hermenêutica enfrenta resistência, como bem alerta Lenio Streck ao denunciar que o processo interpretativo dominante funda-se em um paradigma metafísico, onde o sujeito (jurista ou intérprete) pode dispor dos fatos e dos textos jurídicos de modo livre e consciente (2014).

Em termos de adoção como um processo, metodológico ou filosófico, a depender da abordagem, para determinação da decisão jurídica, a hermenêutica parece ter sido derrotada, tanto na esfera das decisões empreendidas por seres humanos como daquelas tomadas pela inteligência artificial. No primeiro caso, vence as posturas realistas, o pragmatismo, a jurisprudência dos interesses e dos valores. No segundo, as abordagens simbólicas, onde floresceu a hermenêutica computacional, perdem espaço, com o aumento da capacidade computacional, para as abordagens conexionistas. Também é imprescindível reconhecer a impossibilidade de criação de um método de garantia da integridade e correção das decisões judiciais a partir da análise de seu processo de constituição, ou seja, a partir da verificação dos procedimentos utilizados para a sua formação, sejam estas decisões produzidas por um juiz humano ou produzida por métodos computacionais.

Fixa-se, portanto, a necessidade de outra abordagem para a verificação das decisões judiciais, por outro método, que não passe pelo seu processo de criação, mas que se foque no resultado do processo cognitivo, ou seja, na qualidade da fundamentação produzida. Para tanto, antes de abordar a alternativa de utilização de tecnologias conhecidas como “*explainable artificial intelligence*” como forma de produção de decisões fundamentadas, é necessário trazer a este trabalho alguns conceitos importantes como inteligência artificial, hermenêutica computacional, computação cognitiva e aprendizagem de máquina (*machine learning*).

### 3 Inteligência artificial, hermenêutica computacional, computação cognitiva e aprendizagem de máquina (*machine learning*)

Inteligência artificial pode ser conceituada, nos seguintes termos, segundo Bittencour (1996):

O que hoje chamamos Inteligência Artificial (IA) é um ramo da ciência da computação ao mesmo tempo recente, oficialmente nasceu em 1956, e muito antigo, pois a IA foi construída a partir de ideias filosóficas, científicas e tecnológicas herdadas de outras ciências, algumas tão antigas quanto a lógica, com seus 23 séculos.(...)

O objetivo central da IA é simultaneamente teórico – a criação de teorias e modelos para a capacidade cognitiva – e prático – a implementação de sistemas computacionais baseados nestes modelos. Neste sentido, a IA tem uma relação com seu objeto de estudo semelhante a da psicologia, mas com uma importante diferença: os modelos e teorias da IA são implementados em um computador, o que os torna de certa forma autônomos.

Assim, a validade de um modelo ou de uma teoria de IA não precisa ser provada através da comparação de seus resultados previstos com o comportamento psíquico humano, como no caso da psicologia, mas pode ser implementada em um computador e demonstrada diretamente através da ação inteligente do programa no mundo.

Vista a partir da perspectiva teórica, que busca a criação de teorias e modelos para a capacidade cognitiva, não seria estranho a seu próprio objeto, do ponto de vista epistêmico, a aproximação entre inteligência artificial e hermenêutica. Winograd, por exemplo, já na década de 1970, usa insights da hermenêutica filosófica para esboçar uma nova abordagem para o entendimento da linguagem natural (MALLERY; HURWITZ; DUFFY, 1986). Atualmente, é possível observar diferentes abordagens hermenêuticas ao que se pode chamar de hermenêutica digital ou hermenêutica de dados, desde a ideia de uma interpretação e entendimento de textos mediada por computador, até abordagens mais amplas que expandem o espectro de aplicação do texto para conjuntos de dados, onde a informação efetivamente significativa resulta de um processo de interpretação-entendimento que é sempre orientado por pré-juízos (pré-conceitos) tanto do emissor quanto do receptor (ROMELE; SEVERO; FURIA, 2018), em um contexto muito similar ao proposto por Gadamer.

Já computação cognitiva, de modo bem amplo, pode ser conceituada da seguinte forma (S. MODHA et al., 2011):

A computação cognitiva visa desenvolver um mecanismo universal, coerente e unificado, inspirado pelas capacidades da mente. Em vez de montar uma coleção de soluções fragmentadas, em que diferentes processos cognitivos são construídos por meio de soluções independentes, procuramos implementar uma teoria computacional unificada da mente. O pioneiro da IA, Allen Newell, descreveu-o como "um conjunto único de mecanismos para todo o comportamento cognitivo. Nosso objetivo final é uma teoria unificada da cognição humana".<sup>5</sup>

Em uma concepção mais específica, do ponto de vista técnico (KELLY, 2015):

Computação cognitiva refere-se a sistemas que aprendem em escala, raciocinam com propósito e interagem com os humanos naturalmente. Em vez de serem explicitamente programados, eles aprendem e raciocinam a partir de suas interações conosco e de suas experiências com o meio ambiente. Eles são possíveis graças aos avanços em várias áreas científicas ao longo do último meio século, e são diferentes em aspectos importantes dos sistemas de informação que os precederam.

Anteriormente os sistemas eram determinísticos; sistemas cognitivos são probabilísticos. Eles geram não apenas respostas a problemas numéricos, mas também hipóteses, argumentos fundamentados e recomendações sobre conjuntos de dados mais complexos e significativos.<sup>6</sup>

É possível considerar, portanto, a computação cognitiva como uma área da inteligência artificial, cujo método de abordagem para a tomada de decisão funda-se na capacidade da

<sup>5</sup> No original: "Cognitive computing aims to develop a coherent, unified, universal mechanism inspired by the mind's capabilities. Rather than assemble a collection of piecemeal solutions, whereby different cognitive processes are each constructed via independent solutions, we seek to implement a unified computational theory of the mind. AI pioneer Allen Newell described it as 'a single set of mechanisms for all of cognitive behavior. Our ultimate goal is a unified theory of human cognition.'" (tradução livre)

<sup>6</sup> No original: "Cognitive computing refers to systems that learn at scale, reason with purpose and interact with humans naturally. Rather than being explicitly programmed, they learn and reason from their interactions with us and from their experiences with their environment. They are made possible by advances in a number of scientific fields over the past half-century, and are different in important ways from the information systems that preceded them. Those systems have been deterministic; cognitive systems are probabilistic. They generate not just answers to numerical problems, but hypotheses, reasoned arguments and recommendations about more complex — and meaningful — bodies of data." (tradução livre)

aplicação em criar um processo de cognição a partir do conjunto de informações ao qual a aplicação tem acesso, utilizando uma abordagem probabilística. Importante notar, neste ponto, o afastamento entre o método utilizado para obtenção de determinado resultado e o conceito de hermenêutica, em especial da hermenêutica filosófica, de matiz gadameriana.

Mesmo podendo estabelecer uma relação hierárquica, muitas vezes os termos inteligência artificial e computação cognitiva são utilizados como sinônimos, como é o caso, por exemplo, da brilhante explicação dos efeitos da computação cognitiva aplicada aos dados pessoais, feita por Cass Sunstein em seu livro *#Republic* (2017):

Com o crescimento da inteligência artificial, os algoritmos melhorarão imensamente. Eles aprenderão muito sobre você, e eles saberão o que você quer ou vai gostar antes de você e melhor do que você. Eles vão mesmo conhecer suas emoções, novamente antes e melhor do que você, e eles serão capazes de imitar as emoções por conta própria.

(...) O aprendizado da máquina [machine learning] pode ser usado (e provavelmente está sendo usado) para produzir distinções finas. É fácil imaginar uma grande quantidade de classificação, não apenas entre direita e esquerda no âmbito político, mas também com detalhes sobre os assuntos que mais lhe interessa e suas visões prováveis sobre essas questões (imigração, segurança nacional, igualdade, e meio ambiente). Para dizer o mínimo, estas informações poderão ser úteis para outras pessoas - gerentes de marketing, anunciantes, captadores de recursos e mentirosos, incluindo extremistas políticos.<sup>7</sup>

Neste sentido, pode-se fazer uma associação clara entre a computação cognitiva e a capacidade de uma aplicação “aprender” a partir do conjunto de dados aos quais tem acesso, fazendo com que seja possível o estabelecimento de conclusões tendo como base um imenso volume de dados, numa quantidade, muitas vezes, impossível de ser processada por um ser humano. Assim, como bem destaca Kelly, “a meta da computação cognitiva é iluminar aspectos do nosso mundo que anteriormente eram invisíveis – padrões e *insight* em dados não estruturados, em particular – permitindo-nos tomar decisões mais informadas acerca de mais questões consequentes”<sup>8</sup>.

Outra característica fundamental da computação cognitiva está relacionada à operação em linguagem natural. As aplicações baseadas nestas tecnologias conseguem processar dados não estruturados, ou seja, dados que não obedecem a padrões estruturais específicos, como aqueles existentes em bancos de dados convencionais, de forma a obter uma resposta também em linguagem natural. Assim, têm-se como resultado, dentre outras possibilidades, sentenças que fazem sentido para um ser humano.

Assim, não apenas por conseguir processar grandes volumes de dados não estruturados, mas também porque o resultado do processamento pode gerar discursos estruturados, é possível aplicar tais tecnologias para a elaboração de decisões judiciais finais, ou seja, passa a ser viável imaginar aplicações baseadas em computação cognitiva capazes de formular uma sentença a partir da análise de um processo, fundamentando suas decisões.

---

<sup>7</sup> No original: “With the rise of artificial intelligence, algorithms are bound to improve immeasurably. They will learn a great deal about you, and they will know what you want or will like, before you do, and better than you do. They will even know your emotions, again before and better than you do, and they will be able to mimic emotions on their own.(...) Machine learning can be used (and probably is being used) to produce fine-grained distinctions. It is easy to imagine a great deal of sorting— not just from the political right to the political left, but also with specifics about the issues that you care most about, and your likely views on those issues (immigration, national security, equality, and the environment). To say the least, this information can be useful to others — campaign managers, advertisers, fund-raisers and liars, including political extremists.” (tradução livre)

<sup>8</sup> No original: “the goal of cognitive computing is to illuminate aspects of our world that were previously invisible — patterns and insight in unstructured data, in particular — allowing us to make more informed decisions about more consequential matters.” (tradução livre)

As características destas aplicações, capazes de fundamentar suas próprias decisões, são abordadas na próxima seção deste trabalho a partir do conceito de EXplainable Artificial Intelligence (XAI).

#### 4 EXplainable Artificial Intelligence (XAI) e as decisões computacionalmente fundamentadas

Aplicações baseadas em inteligência artificial podem gerar decisões injustas? Esta é a pergunta central que se coloca ao analisar a viabilidade de uso destas aplicações no âmbito jurídico. Um sinal de alerta foi ligado, em 2016, quando o jornal americano *The New York Times*, publicou a matéria “Inteligência artificial: problema dos ‘caras brancos’”<sup>9</sup>, onde se denuncia, a partir de estudo realizado pela entidade de jornalismo investigativo ProPublica, a existências de vieses de decisão em sistemas de inteligência artificial utilizados na análise de previsão de reincidência de criminosos, com maior incidência de falsos positivos para pessoas negras e maior incidência de falsos negativos para pessoas brancas.

Na matéria, a jornalista Kate Crawford (2016) comenta sobre o desconhecimento dos motivos que levaram à existência de um viés étnico, nos seguintes termos:

A razão pela qual essas previsões são tão distorcidas ainda é desconhecida, porque a empresa responsável por esses algoritmos mantém suas fórmulas em segredo - é uma informação proprietária. Os juízes se baseiam em avaliações de risco conduzidas por máquinas de diferentes formas - alguns podem até mesmo desconsiderá-las totalmente -, mas pouco podem fazer para entender a lógica por trás delas.<sup>10</sup>

É fácil notar o desconforto em basear uma decisão relativa à liberdade em um resultado único, sem que se tenha nenhuma informação quanto ao método utilizado, do conjunto de dados analisados e das relações estabelecidas entre estes dados a partir do método aplicado. Tem-se apenas um número, relativo à probabilidade de reincidência sem nenhuma fundamentação.

O grande impacto das decisões computacionalmente construídas com potencial de restrição do exercício de direitos fundamentais levou ao aumento da consciência quanto à necessidade de se desenvolver, tecnicamente, sistemas de inteligência artificial capazes de justificar o resultado obtido pela aplicação, em uma linguagem que possa ser plenamente entendida por um ser humano.

Paradoxalmente, a DARPA (*Defense Advanced Research Projects Agency*), ligado ao Departamento de Defesa dos EUA, anunciou em agosto de 2017 (GORDON-MURNANE, 2018), a criação do programa *Explainable Artificial Intelligence*, que segundo o Dr. David Gunning, gerente responsável, “será essencial para que os futuros combatentes compreendam, confiem plenamente e gerenciem efetivamente uma geração de máquinas de apoio baseadas em inteligência artificial”<sup>11</sup> (GUNNING, [s.d.]). O mesmo autor segue explicando, resumidamente, o modo de funcionamento desta nova arquitetura de sistemas de aprendizado de máquina:

<sup>9</sup> No original: “Artificial Intelligence’s white guy problem.” (tradução livre)

<sup>10</sup> No original: “The reason those predictions are so skewed is still unknown, because the company responsible for these algorithms keeps its formulas secret — it’s proprietary information. Judges do rely on machine-driven risk assessments in different ways — some may even discount them entirely — but there is little they can do to understand the logic behind them.” (tradução livre)

<sup>11</sup> No original: “will be essential if future warfighters are to understand, appropriately trust, and effectively manage an emerging generation of artificially intelligent machine partners.” (tradução livre)

Os novos sistemas de aprendizado de máquina terão a capacidade de explicar seu ‘modo de pensar’, caracterizar seus pontos fortes e fracos e transmitir uma compreensão sobre como se comportarão no futuro. A estratégia para alcançar esse objetivo é desenvolver técnicas novas ou modificar técnicas existentes de aprendizado de máquina que produzirão modelos mais auto-explicáveis. Esses modelos serão combinados com técnicas de interface homem-computador de última geração capazes de traduzir os modelos em diálogos com explicações compreensíveis e úteis para o usuário final.<sup>12</sup>

Considerando a expansão do uso da inteligência artificial, mais especificamente do *machine learning*, em sistemas críticos de diversas áreas como, por exemplo, diagnósticos médicos, concessão de crédito, apoio a decisão jurídica, sistemas militares, controle aéreo, dentre outras, torna-se fundamental que as decisões sugeridas (ou tomadas) por estes sistemas possam ser validadas (ou auditadas) por seres humanos, tornando-se necessária a implementação de modelos computacionais capazes de abrir a caixa preta, expondo, em linguagem natural ou visualmente, os fundamentos que justificam a decisão tomada pela aplicação (SAMEK; WIEGAND; MÜLLER, 2017).

Apesar da origem militar, a ideia de se estabelecer um modelo de sistema de aprendizado de máquina que possa, além de gerar um resultado válido, também prover uma fundamentação adequada de suas conclusões possui absoluta aderência à necessidade de estabelecimento de controle sobre decisões computacionais no âmbito do Direito.

Passa a fazer sentido, a partir deste conceito, a concepção de verdadeiros sistemas de decisão judicial capazes de promover a fundamentação em um determinado padrão esperado, garantindo às partes e, também, aos outros atores processuais, o pleno entendimento da decisão em padrões conhecidos e pré-determinados.

Mesmo ainda sendo uma realidade distante, principalmente para situações que envolvam casos difíceis, a ideia de se criar sistemas de aprendizagem de máquina que, por concepção de projeto, possui a capacidade de fundamentar adequadamente as decisões tomadas segundo critérios específicos, aumenta as expectativas pela criação de aplicações que possam aumentar de modo substancial a eficácia do processo de decisão processual, tanto no âmbito judicial quanto na seara administrativa, sem que para tanto seja necessário sacrificar a fundamentação das decisões, verdadeiro corolário do devido processo legal.

## 5 Conclusão

A primeira conclusão que se pretende extrair deste trabalho diz respeito à proximidade entre os processos de decisão humano e computacional no que tange à forma como as decisões são concebidas. Não há como estabelecer um comportamento ideal, definido a priori, para o juiz, assim como não é possível garantir um modelo infalível de decisão computacional. Esta aproximação é facilmente constatada ao se observar que o resultado final obtido a partir do trabalho cognitivo, tanto do juiz, quanto do programador, estão sujeitos a influência de fatores externos à sua atividade, acarretando um certo grau de incerteza quanto ao resultado final obtido e, conseqüentemente, também um certo grau de desconfiança quanto a este mesmo resultado.

---

<sup>12</sup> No original: “New machine-learning systems will have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future. The strategy for achieving that goal is to develop new or modified machine-learning techniques that will produce more explainable models. These models will be combined with state-of-the-art human-computer interface techniques capable of translating models into understandable and useful explanation dialogues for the end user.” (tradução livre)

Feita esta constatação fica evidente que o método de controle da eficácia e qualidade das decisões judiciais computacionalmente elaboradas passa, essencialmente, pela garantia da devida fundamentação destas decisões pelo próprio sistema que a gerou. Alinha-se, desta forma, o sistema de controle de decisões humana e computacional sobre os mesmos parâmetros. Como bem destaca Streck (2014), “[a] peculiaridade do direito reside no fato de que aquele que interpreta precisa, necessariamente, dar à comunidade política as razões de sua interpretação”.

Por fim, os avanços das técnicas computacionais, que visam a inclusão de novo processo de aprendizagem de máquina dedicado à fundamentação das decisões, em uma linguagem adequada, que possibilite o entendimento das decisões tomadas por seres humanos, o efetivo controle das decisões e, conseqüentemente, o aumento do nível de confiança das pessoas nestes sistemas de decisão computacional são capazes de criar um cenário promissor para o desenvolvimento de sistemas de decisões judiciais fundamentadas computacionalmente, o que certamente contribuiria para o aumento expressivo da efetividade da jurisdição, principalmente em um país de alto nível de judicialização como é o caso do Brasil.

## 6 Referências

- BITTENCOURT, G. **Inteligência Artificial. Ferramentas e Teorias**. Campinas: Instituto de Computação. UNICAMP, 1996.
- CRAWFORD, K. **Artificial Intelligence’s White Guy Problem**. 2016. Disponível em: <<https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>>. Acesso em: 8 maio. 2018.
- CURY, C. **Caminho para agilizar a Justiça**. 2015. Disponível em: <<https://oglobo.globo.com/opiniao/caminho-para-agilizar-justica-17289105>>. Acesso em: 7 maio. 2018.
- GORDON-MURNANE, L. Ethical, Explainable Artificial Intelligence: Bias and Principles. **Online Searcher**, [s. l.], v. 42, n. 2, p. 22–44, 2018. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=128582745&lang=pt-br&site=ehost-live>>
- GUNNING, D. **Explainable Artificial Intelligence**. [s.d.]. Disponível em: <<https://www.darpa.mil/program/explainable-artificial-intelligence>>. Acesso em: 8 maio. 2018.
- KELLY, J. E. Computing, cognition and the future of knowing. **IBM Global Services**, [s. l.], p. 7, 2015.
- MALLERY, J. C.; HURWITZ, R.; DUFFY, G. Hermeneutics: From Textual Explication to Computer Understanding? **Massachusetts Institute of Technology. Artificial Intelligence Laboratory**, Cambridge, 1986.
- MARANHÃO, J. O impacto na Justiça. **Exame CEO**, São Paulo, 2018.
- MELLO, P. P. C. **Nos bastidores do STF**. Rio de Janeiro: Forense, 2015.
- ORDEM DOS ADVOGADOS DO BRASIL - SECCIONAL RIO DE JANEIRO. **OAB/RJ convoca ato contra “Justiça” sem advocacia – OAB/RJ**. 2018. Disponível em: <<http://www.oabrj.org.br/noticia/111704-oabrj-convoca-ato-contrajustica-semadvocacia>>. Acesso em: 7 maio. 2018.
- PÁDUA, L. **OAB-RJ protesta contra sistema automatizado de solução de conflitos**. 2018. Disponível em: <<https://www.jota.info/justica/oab-rj-protesta-contrasistema-automatizado-de-solucao-de-conflitos-19032018>>. Acesso em: 7 maio. 2018.
- POSNER, R. A. **How judges think**. Cambridge, Mass.: Harvard University Press, 2008.
- PUPO, F.; OLIVON, B. **Receita testa inteligência artificial em julgamentos**. 2018. Disponível em: <<http://www.valor.com.br/legislacao/5473055/receita-testa-inteligencia-artificial-em-julgamentos>>. Acesso em: 7 maio. 2018.
- ROMELE, A.; SEVERO, M.; FURIA, P. Digital hermeneutics: from interpreting with machines to interpretational machines. **Ai & Society**, [s. l.], v. 0, n. 0, p. 0, 2018. Disponível em: <<http://link.springer.com/10.1007/s00146-018-0856-2>>
- S. MODHA, D.; ANANTHANARAYANAN, R.; ESSER, S. K.; NDIRANGO, A.; SHERBONDY, A.;

CAMARGO, Gustavo Xavier de. Decisões judiciais computacionalmente fundamentadas: uma abordagem a partir do conceito de explainable artificial intelligence. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 167-177, 2019.

- SINGH, R. Cognitive Computing. **Communications of the ACM**, [s. l.], v. 54, n. 8, p. 62–71, 2011.
- SAMEK, W.; WIEGAND, T.; MÜLLER, K.-R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. **CoRR**, [s. l.], 2017.
- STRECK, L. **Hermenêutica Jurídica e(m) Crse**. 11ª Ed. ed. Porto Alegre: Livraria do Advogado Editora, 2014.
- SUNSTEIN, C. **#Republic: divided democracy in age of social media**. Princeton: Princeton University Press, 2017.
- THOMPSON, K. Reflections on trusting trust. **Communications of the ACM**, New York, v. 27, n. 8, p. 761–763, 1984. Disponível em: <<http://portal.acm.org/citation.cfm?doid=358198.358210>>