

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS JURÍDICAS
CURSO DE GRADUAÇÃO EM DIREITO

DANIEL HENRIQUE ARRUDA BOEING

**ENSINANDO UM ROBÔ A JULGAR: PRAGMÁTICA, DISCRICIONARIEDADE E
VIESES NO USO DE APRENDIZADO DE MÁQUINA NO JUDICIÁRIO**

FLORIANÓPOLIS

2019

DANIEL HENRIQUE ARRUDA BOEING

**ENSINANDO UM ROBÔ A JULGAR: PRAGMÁTICA, DISCRICIONARIEDADE E
VIESES NO USO DE APRENDIZADO DE MÁQUINA NO JUDICIÁRIO**

Trabalho de Conclusão de Curso de Graduação
em Direito do Centro de Ciências Jurídicas da
Universidade Federal de Santa Catarina como
requisito para a obtenção do título de Bacharel em
Direito.

Orientador: Prof. Dr. Alexandre Morais da Rosa

FLORIANÓPOLIS

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Boeing, Daniel Henrique

Ensinando um robô a julgar: pragmática,
discricionariiedade e vieses no uso de aprendizado
do máquina no judiciário / Daniel Henrique Boeing ;
orientador, Alexandre Morais da Rosa, 2019.

85 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de
Ciências Jurídicas, Graduação em Direito,
Florianópolis, 2019.

Inclui referências.

1. Direito. 2. Teoria do direito. 3. Aprendizado
de máquina. 4. Filosofia da linguagem. 5.
Pragmática, vieses e discricionariiedade. I. Rosa,
Alexandre Morais da. II. Universidade Federal de
Santa Catarina. Graduação em Direito. III. Título.

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS JURÍDICAS
COLEGIADO DO CURSO DE GRADUAÇÃO EM DIREITO

TERMO DE APROVAÇÃO

O presente Trabalho de Conclusão de Curso, intitulado "Ensinando um robô a julgar: pragmática, discricionariedade e vieses no uso de aprendizado de máquina no judiciário", elaborado pelo acadêmico "Daniel Henrique Arruda Boeing", defendido em 02/12/2019 e aprovado pela Banca Examinadora composta pelos membros abaixo assinados, obteve aprovação com nota 10 (dez), cumprindo o requisito legal previsto no art. 10 da Resolução nº 09/2004/CES/CNE, regulamentado pela Universidade Federal de Santa Catarina, através da Resolução nº 01/CCGD/CCJ/2014.

Florianópolis, 02 de dezembro de 2019



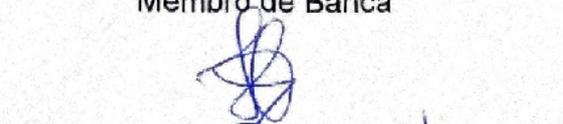
Alexandre Moraes da Rosa
Professor Orientador



Marília Raposo Vieira
Membro de Banca



Luana Renostro Heinen
Membro de Banca



Fernando Pacheco Amorim
Membro de Banca

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS JURÍDICAS
COLEGIADO DO CURSO DE GRADUAÇÃO EM DIREITO

TERMO DE RESPONSABILIDADE PELO INEDITISMO DO TCC E
ORIENTAÇÃO IDEOLÓGICA

Aluno(a): Daniel Henrique Arruda Boeing

RG: 10.133.794-4

CPF: 108.449.399-35

Matrícula: 14201231

Título do TCC: "Ensinando um robô a julgar: pragmática, discricionariedade e vieses no uso de aprendizado de máquina no judiciário"

Orientador(a): Prof. Dr. Alexandre Moraes da Rosa

Eu, Daniel Henrique Arruda Boeing, acima qualificado, venho, pelo presente termo, assumir integral responsabilidade pela originalidade e conteúdo ideológico apresentado no TCC de minha autoria, acima referido.

Florianópolis, 02 de dezembro de 2019

Daniel Boeing

Daniel Henrique Arruda Boeing

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS JURÍDICAS
COLEGIADO DO CURSO DE GRADUAÇÃO EM DIREITO

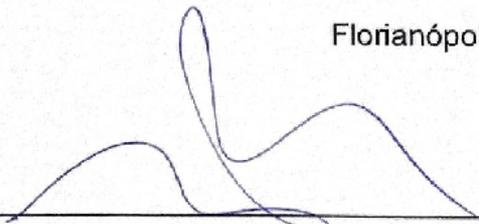
ATA DE SESSÃO DE DEFESA DE TCC

Aos dois dias do mês de dezembro do ano de 2019, às 20 horas, na Sala 109 do CCJ, foi realizada a defesa pública do Trabalho de Conclusão de Curso (TCC) intitulado "Ensinando um robô a julgar: pragmática, discricionariedade e vieses no uso de aprendizado de máquina no judiciário", elaborado pelo(a) acadêmico(a) Daniel Henrique Arruda Boeing, matrícula nº 14201231, composta pelos membros Alexandre Morais da Rosa, Luana Renostro Heinen e Marília Raposo Vieira, abaixo assinados, obteve a aprovação com nota 10 (DT), cumprindo o requisito legal previsto no art. 10 da Resolução nº 09/2004/CES/CNE, regulamentado pela Universidade Federal de Santa Catarina, através da Resolução nº 01/CCGD/CCJ/2014.

Aprovação Integral

Aprovação Condicionada aos seguintes reparos, sob fiscalização do Prof. Orientador

Florianópolis, 02 de dezembro de 2019



Alexandre Morais da Rosa

Professor Orientador



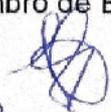
Marília Raposo Vieira

Membro de Banca



Luana Renostro Heinen

Membro de Banca


Fernando Pacheco Amorim

Membro de Banca

O homem que diz: Dou - Não dá!

Porque quem dá mesmo

Não diz!

O homem que diz: Vou - Não vai!

Porque quando foi já

Não quis!

O homem que diz sou - Não é!

Porque quem é mesmo é

Não sou!

O homem que diz: Tô - Não tá!

Porque ninguém tá

Quando quer

Coitado do homem que cai

No canto de Ossanha, traidor!

Coitado do homem que vai

Atrás de mandinga de amor

Vai! Vai! Vai! Vai! - Não Vou!

Que eu não sou ninguém de ir

Em conversa de esquecer

A tristeza de um amor

Que passou

Não!

Eu só vou se for pra ver

Uma estrela aparecer

Na manhã de um novo amor

(Canto de Ossanha - Vinicius de Moraes)

RESUMO

Este Trabalho de conclusão de curso explora algumas das questões relativas à implementação de ferramentas de aprendizado de máquina na administração da justiça. Em um primeiro momento, o trabalho discute as peculiaridades do aprendizado de máquina em face de outros modelos estatísticos, bem como quais são os requisitos necessários para que tais algoritmos possam ser utilizados na performance de argumentação jurídica (*legal reasoning*). Em seguida, resgatam-se algumas das discussões da teoria do direito e demonstra-se como elas estão relacionadas à filosofia da linguagem. Defende-se que tanto Hans Kelsen, quanto Herbert Hart, ainda que partam de caminhos diferentes, chegam ao problema do decisionismo e que isso é uma consequência do paradigma por eles adotado, a saber, a filosofia analítica. Adiante, discute-se como linguagens artificiais são capazes de assimilar o âmbito pragmático de linguagens naturais e de que forma processos decisórios humanos e algorítmicos são afetados por vieses. Por fim, elencam-se três tipos de uso mais recorrentes do aprendizado de máquina no judiciário.

Palavras-chave: Aprendizado de máquina. Teoria do direito. Filosofia da linguagem. Pragmática. Vieses. Decisionismo. Hans Kelsen. Herbert Hart. Ludwig Wittgenstein.

ABSTRACT

This final paper goes on some of the issues surrounding the usage of machine learning tools in the administration of justice. At first, are brought some of the peculiarities of machine learning in the face of other statistical models, as well as the prerequisites for such algorithms to perform legal reasoning. Then, some of the discussions of the theory of law are rescued and it is demonstrated how they relate to the philosophy of language. It is argued that both Hans Kelsen and Herbert Hart, though starting from different outsets, come to the problem of decisionism and that this is a consequence of their elected paradigm, namely the analytic philosophy. Ahead, it is discussed how artificial languages are able to assimilate the pragmatic dimension of natural languages and how human and algorithmic decision-making processes are affected by heuristics and biases. Finally, it is presented three most recurring ways of implementation of machine learning into the judiciary.

Key words: Machine learning. Theory of law. Philosophy of language. Pragmatics. Biases. Decisionism. Hans Kelsen. Herbert Hart. Ludwig Wittgenstein.

SUMÁRIO

Introdução	11
Capítulo 1 - Artificialmente inteligente: como máquinas “aprendem”	16
1.1 O que é aprendizado de máquina?	16
1.1.1 Agentes inteligentes	17
1.1.2 Como máquinas “aprendem”: peculiaridades do machine learning	21
1.2 Aplicando aprendizado de máquina a textos legais	24
1.2.1 Especificidades do campo jurídico	24
1.2.2 Vetorizando textos	27
Capítulo 2 - A linguagem no direito: pragmática e discricionariedade	33
2.1 O positivismo científico e a Teoria Pura do Direito de Kelsen: a visão descritivista da linguagem	35
2.2 Jogos de linguagem e o Conceito de Direito de Hart: a virada pragmática	40
2.3 O decisionismo dentro do paradigma da filosofia analítica	45
Capítulo 3 - Artificialmente parcial: contexto e vieses em algoritmos	51
3.1 Ensinando a máquinas o contexto	52
3.2 Vieses: como opiniões se tornam “ciência”	57
3.2.1 Heurística e vieses	57
3.2.2 Vieses em máquinas	61
3.3 Formas de se utilizar o aprendizado de máquina no judiciário	67
3.3.1 A questão da corrobótica	68
3.3.2 Três tipos de uso do aprendizado de máquina no direito	71
Conclusão	78
Referências	81

Introdução

Tudo pode ser traduzido em termos de informação. Desde *spins* de elétrons, em uma escala subatômica, a ondas gravitacionais definindo as órbitas de corpos celestes, ou o código genético nas células de seres vivos, até as palavras que formam este texto, tudo carrega consigo, de um jeito ou de outro, informação. Linguagens nada mais são que maneiras de propagar informação, processo que pode ocorrer de forma imediata ou pelo intermédio de um outro objeto, diferente daqueles que se comunicam. Neste último caso, diz-se que houve a atuação de um signo, um sinal qualquer que evoca um certo objeto (real ou não) a um determinado intérprete, estabelecendo uma relação de significação.

O processo por meio do qual algo funciona como signo é chamado de semióse, principal objeto de estudo da semiótica.¹ Sendo ou não um feito exclusivamente humano, o processo semiótico adquiriu vital importância nesta espécie, que passou a utilizá-lo das mais variadas formas. Um simples gesto, uma palavra, escrita ou falada, ou zeros e uns processados em um computador permitem veicular ideias capazes de manipular comportamentos e estabelecer relações de cooperação (ou disputa) entre diferentes indivíduos.

Ainda assim, o processo comunicativo humano é um refém dos signos. Sem o intermédio destes, não há como transmitir pensamentos de um indivíduo a outro, de forma que se torna de especial importância buscar entender de que forma se dá a relação de dependência entre o comunicar humano e a ação dos signos. Ignorar, entretanto, o caráter por vezes ambivalente e dúbio dessa mediação é negar o próprio movimento da linguagem. É sob essa ótica que a questão da implementação, nos tribunais, de algoritmos de aprendizado de máquina, um dos subcampos da inteligência artificial (IA), será analisada.

Avanços recentes na área da ciência da computação tiveram reflexos em diversos segmentos sociais. Alterou-se radicalmente a forma como hoje se dão interações sociais, econômicas e governamentais e na base da maior parte dessas transformações está a disseminação de técnicas de IA, em especial, de aprendizado

¹ Cf. SANTAELLA, Lúcia. O que é semiótica. São Paulo: Editora Brasiliense, 1983.

de máquina, por meio das quais foram possíveis avanços sem precedentes na análise e gestão de grandes quantidades de informações.

O campo do direito já é objeto de aplicação de tais inovações, o que, ao que tudo indica, é um processo que tende a se potencializar, dada a carga de trabalho dos tribunais e a natureza das atividades que eles exercem. Somente em 2018, chegaram ao judiciário brasileiro mais de 28 milhões de casos novos e cada juiz julgou, em média, 1877 processos (quase 8 por dia útil), sendo que ainda restam quase 79 milhões de casos pendentes de julgamento, dos quais 39% são execuções fiscais.²

Dado esse cenário, é fácil conceber como o poder judiciário se torna um ambiente bastante fértil para o implemento de soluções inovadoras, o que se traduz em diversas tentativas, já em andamento, de utilização de IAs por tribunais espalhados pelo país³. Ainda assim, são relativamente escassos, no campo jurídico, debates sobre as consequências da transição que está em vias de ocorrer.

Haja vista a incipiência de tais mudanças, torna-se difícil prever com precisão quais serão suas consequências, que podem ser diversas. Por um lado, o uso de IAs promete maior acesso à justiça e eficiência em diferentes tipos de tarefas, por outro, pode-se estar caminhando em direção a um cenário em que algoritmos enviesados, envoltos pelo manto da imparcialidade e cientificidade, decidem o futuro das pessoas em aspectos sensíveis de suas vidas.

Para garantir que tais mudanças ocorram de forma compatível com os princípios basilares de um Estado democrático e de direito, é necessário, desde já, estabelecer-se quais objetivos que se quer alcançar com as IAs, mas, sobretudo, o que se espera evitar de um tal uso. Nesse sentido, o papel e a postura dos juristas serão fundamentais nessa transição, pois eles atuarão tanto na regulamentação do setor, quanto na própria administração da justiça, que, daqui não muito tempo, será provavelmente inconcebível sem algum tipo de uso de IA.

² BRASIL. CONSELHO NACIONAL DE JUSTIÇA. Justiça em Números 2019: Relatório Analítico, pp. 36, 37, 89 e 131.

³ Para citar alguns exemplos: Radar, no Tribunal de Justiça de Minas Gerais, Victoria, no Tribunal de Justiça do Rio de Janeiro, Victor, no Supremo Tribunal Federal e Bem-Te-Vi, no Tribunal Superior do Trabalho.

O objetivo deste trabalho é, dessa forma, propiciar um entendimento, em linhas gerais, do modo de funcionamento de IAs, mais especificamente, do processamento de linguagem natural (NLP), e qual sua relação com a prática jurídica. Entender como eles operam propiciará também a compreensão de parte de seus limites, o que é imprescindível para sua adequada utilização.

É claro que não se almeja dar uma resposta definitiva ao problema. A uma, porque a tecnologia está em constante mudança, de forma que quanto mais se aprofunda em uma abordagem, maior risco se corre de cair na obsolescência. A duas, pois a própria natureza das questões envolvidas, que dizem respeito a fenômenos sociais complexos, não se mostra compatível com respostas que se pretendam “finais”.

Trata-se, na verdade, de um trabalho que busca apresentar uma das possíveis formas de se encarar a questão do direito e da inteligência artificial e como esses campos se entrelaçam por meio da linguagem humana. Para tanto, em um primeiro momento são abordados conceitos iniciais sobre o aprendizado de máquina e algumas das especificidades de sua utilização no campo jurídico. Em seguida, no segundo capítulo, demonstra-se como debates da teoria do direito estão intimamente vinculados à filosofia da linguagem, de modo que diferentes formas de entendê-la resultam em diferentes visões sobre o fenômeno jurídico.

O terceiro e último capítulo traz algumas preocupações da filosofia da linguagem ao processamento de linguagem natural, a saber, como máquinas são capazes de assimilar o contexto. Feitas tais considerações, passa-se ao problema de como vieses afetam tanto decisões humanas, quanto algoritmos, e o que se deve fazer para mitigar sua ocorrência nestes últimos. Por fim, elencam-se três principais tipos de uso do aprendizado de máquina no judiciário.

Antes de se prosseguir, é necessário, contudo, levar em conta que as questões sobre filosofia da linguagem aqui exploradas se inserem, quase que integralmente, nos pressupostos do paradigma da filosofia analítica. Adota-se o marco teórico proposto por Júlio Cabrera⁴, segundo o qual as diferentes filosofias da linguagem possuem quatro “tendências” ou abordagens: analíticas, hermenêuticas, fenomenológicas e metacríticas.

⁴ Cf. CABRERA, Julio. Margens das Filosofias da Linguagem.

Todas têm em comum o fato de não considerarem a linguagem como um mero veículo de transmissão do pensamento (concepção veicular), mas ela própria como essencial na constituição dos significados (concepção constitucional). Ainda assim, diferentes abordagens trarão diferentes preocupações, que nem sempre poderão ser explicadas ou não farão sentido aos conceitos de outros modos de se pensar a linguagem.

As filosofias analíticas se caracterizam por excluir os elementos vividos, vez que estes são “não decomponíveis e não articuláveis” e, assim, por privilegiar o espectro mais “lógico-matemático” da linguagem e sua função cognitiva. No outro extremo da linguagem, encontram-se as filosofias hermenêuticas, entendidas como “qualquer filosofia que coloque a linguagem em contato com elementos experienciais e atribua à história e ao vivido dimensão de significatividade”⁵. Nesse sentido, hermenêutica e análise da linguagem são contradições em termos. Três pontos principais explicam o contraste entre os modos do pensar filosófico: i) tipo de experiência que cada uma delas leva em consideração; ii) acentos dados às funções da linguagem; iii) pressupostos de compreensão: existência ou não de estruturas *a priori* de compreensão.⁶

A fenomenologia da linguagem, partindo da concepção fenomenológica de E. Husserl, insere-se em uma “filosofia geral da intencionalidade da consciência, dentro da qual a linguagem encontra sua função e seus limites”⁷. Na fenomenologia, os “sentidos”, saem da esfera da linguagem e passam a ser vinculados a atos intencionais, de forma que os signos (sinais sonoros, visuais, gesticulativos, etc.) adquirem significação somente quando ligados à intencionalidade expressiva, uma forma especial de intencionalidade.

Filosofias metacríticas buscam “distorções” básicas da linguagem que não são captadas por conceitos analíticos, hermenêuticos ou fenomenológicos. Para tanto, busca-se nos signos traços daquilo que é sistematicamente escondido ou distorcido.⁸ A metacrítica se caracteriza pela recusa ao entendimento, mesmo quando o discurso explícito, à primeira vista, se mostra bem formulado. Exemplos

⁵ CABRERA, op. cit., p. 29.

⁶ CABRERA, op. cit., pp. 31-39.

⁷ CABRERA, op. cit., p. 93.

⁸ CABRERA, op. cit., p. 172.

desse modo de pensar são encontrados nas obras de Karl Marx e Sigmund Freud, vez que elas lidam, respectivamente, com o discurso ideológico e perturbações na fala.

Capítulo 1 - Artificialmente inteligente: como máquinas “aprendem”

1.1 O que é aprendizado de máquina?

Machine learning ou aprendizado de máquina não é uma novidade. O termo existe desde 1959 e é usado para referir-se a algoritmos que podem “aprender” a partir de dados e fazer previsões⁹. Foi apenas recentemente, contudo, que abordagens mais aprimoradas, tais como o *deep learning*, se tornaram possíveis, graças aos avanços tecnológicos e da imensa quantidade de dados disponíveis na era do *big data*, motivo pelo qual seu uso vem ganhando especial atenção de empresas e governos.

Nas próximas seções, além de serem feitas as precisões terminológicas relevantes, será explicado como máquinas “aprendem” e como tal processo se distingue de outros métodos para realizar medições e previsões. Uma vez compreendidas as peculiaridades desses modelos, poderão ser analisados com maior precisão os seus diversos tipos de uso.

Antes, contudo, faz-se necessário definir os termos “algoritmo” e “modelo”, que serão bastante utilizados ao longo deste trabalho. Um algoritmo nada mais é que um conjunto finito e preciso de passos para resolver um problema ou responder uma questão.¹⁰ Assim, uma receita de bolo consiste em um algoritmo, escrito em linguagem natural, para o problema “como fazer um bolo”. Em ciência da computação, usualmente, se denominam “algoritmos” programas de computador escritos em linguagem de programação para executar diferentes variedades de tarefas.

Por modelo, entende-se “uma estrutura que sumariza padrões de dados de maneira estatística ou lógica, de forma que ele pode ser aplicado em novos dados”¹¹.

⁹ SAMUEL, Arthur. *Some Studies in Machine Learning Using the Game of Checkers*, pp. 210–229.

¹⁰ Cf. The Editors Of Encyclopaedia Britannica. Algorithm: “systematic procedure that produces—in a finite number of steps—the answer to a question or the solution of a problem”. Disponível em: <<https://www.britannica.com/science/algorithm>>. Acesso em 19 nov. /2019.

¹¹ Tradução livre: “a model is a structure that summarizes the patterns in data in some statistical or logical form in which it can be applied to new data” em ASHLEY, Kevin D.. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, p. 234. Um outra definição de modelo será explorada no Capítulo 3, deste trabalho, formulada por Cathy O’Neil e mais voltada para a questão dos vieses.

A diferença entre um modelo e um algoritmo diz respeito essencialmente a dados. Enquanto um algoritmo é um método ou procedimento abstrato, o modelo é o resultado da utilização de um algoritmo em um conjunto específico de dados, por meio do qual se transforma valores de entrada (*inputs*) em valores de saída (*outputs*), procedimento que pode ser aplicado a novos dados para fazer previsões.

1.1.1 Agentes inteligentes

Para se entender o aprendizado de máquina, é necessário explorar conceitos básicos sobre inteligência artificial (IA), cuja definição está longe de ser um ponto pacífico. Um ponto de partida, nessa empreitada, é o Teste de Turing, um experimento hipotético, proposto em 1950, por Alan Turing, que apresenta uma das possíveis abordagens do problema.¹² Nele, máquinas seriam avaliadas de acordo com sua capacidade de mimetizar seres humanos, de forma que o computador passaria no teste se um interrogador humano, após fazer perguntas por escrito, não conseguisse identificar estar se comunicando com outro ser humano ou com um robô.

Para tanto, o computador precisaria, no mínimo, das seguintes habilidades: *processamento de linguagem natural (natural language processing)*: para conseguir se comunicar com os seres humanos através de linguagens naturais (tais como o português); *representação de conhecimento (knowledge representation)*: para armazenar seus conhecimentos; *raciocínio automatizado (automated reasoning)*: para responder questões e chegar a novas conclusões a partir do conhecimento acumulado e; *aprendizado de máquina (machine learning)*: para se adaptar a novas circunstâncias e detectar padrões. Tais disciplinas, junto com a *robótica* e a *visão computacional*, formam boa parte dos campos do que hoje é entendido como inteligência artificial.¹³

Não obstante, pouco esforço foi desenvolvido no sentido de criarem-se máquinas que obtivessem êxito no Teste de Turing. Pesquisadores da área normalmente não restringem seus trabalhos à imitação de características atribuídas

¹² RUSSEL, S., Artificial Intelligence: a modern approach, p.2.

¹³ RUSSEL, S., op. cit., p.2.

ao pensar racional humano, preferindo relacionar IAs ao desenvolvimento de *agentes inteligentes*¹⁴. Opta-se por essa abordagem, pois se entende que a racionalidade humana é voltada a um ambiente bastante específico e decorre de um processo biológico evolutivo. a racionalidade, concebida de maneira abstrata, possui raízes matemáticas bem definidas e pode prescindir de pensamento, não se limitando a tão somente regras de inferência.

Desse modo, a racionalidade “abstrata” pode se desdobrar em outros tipos de racionalidade, a depender dos diferentes contextos em que os agentes estão inseridos.¹⁵ Poderia-se definir, então, o campo de estudos da inteligência artificial como o conjunto de esforços “concentrados em princípios gerais de agentes racionais e nos componentes para construí-los”¹⁶.

Assim, percebem-se algumas das características da IA. Em primeiro lugar, é importante notar que ela não se refere a robôs em si, visto que estes seriam apenas a “carcaça” que abriga os softwares que executam IAs. Em segundo lugar, fica evidente que se trata de um conceito bastante amplo, que abrange desde um algoritmo que recomenda filmes, passando por carros autônomos e uma eventual superinteligência. Nesse sentido, há dois tipos (ou calibres) de IA: forte e fraca.

Uma IA fraca ou de propósito limitado (*Artificial Narrow Intelligence - ANI*) é aquela especializada em uma única área. Assim, ela executa tarefas específicas muito bem (geralmente melhor que os seres humanos), mas não consegue fazer nada além daquela tarefa¹⁷. Exemplos de IA fraca são programas que: jogam xadrez, recomendam vídeos com base no histórico do usuário ou sugestões de produtos em sites de compra, identificam números escritos à mão, preveem preços de ações, reconhecem sons e imagens como palavras ou objetos, dentre outros.

Uma IA forte ou de propósito geral (*Artificial General Intelligence - AGI*), também chamada de Inteligência Artificial de nível humano, refere-se a um computador que consegue desempenhar qualquer tarefa intelectual que um humano seja capaz de fazer.¹⁸ Assim, uma IA de propósito geral busca assemelhar-se à

¹⁴ RUSSEL, S., op. cit., p.2

¹⁵ RUSSEL, S., op. cit., p. 4.

¹⁶ RUSSEL, S., op. cit., p. 5.

¹⁷ STANKOVIC, Mirjana et al. Exploring legal, ethical and policy implications of AI, p. 5.

¹⁸ STANKOVIC, Mirjana et al., op. cit., p. 5.

inteligência humana, o que pressupõe habilidades de: raciocinar, representar conhecimento, planejar, aprender, comunicar-se em linguagem natural e utilizar todos esses atributos conjuntamente para atingir objetivos específicos¹⁹. Importante salientar que, ainda que haja esforços para a criação de programas com escopos que envolvam tarefas de diferentes naturezas, tais como carros autônomos ou o *Watson* da IBM, ainda se está muito aquém das capacidades de abstração, raciocínio e processamento de dados do cérebro humano.

Também bastante abrangente, o aprendizado de máquina ou *machine learning*, que será mais propriamente o objeto de estudo deste trabalho, é, em síntese, uma das maneiras de se obter IA. O termo foi cunhado por Arthur Samuel, cientista pioneiro no campo, que o define como “a habilidade para aprender sem ter sido explicitamente programado”.²⁰

Apesar de fornecer uma ideia geral sobre o assunto, tal definição é um pouco vaga, pois não aborda um dos aspectos essenciais do aprendizado de máquina: melhorar um critério de performance através da experiência. Seu propósito é aprender a partir de dados anteriores, com o objetivo de fazer o maior número possível de previsões corretas em novos dados, ainda não vistos ou classificados.²¹

Classifica-se o aprendizado em supervisionado e não supervisionado. A distinção diz respeito a se o algoritmo foi treinado a partir de dados que foram ou não anteriormente classificados por humanos²². No aprendizado supervisionado, houve uma classificação prévia, enquanto o algoritmo não supervisionado procura por estruturas semelhantes dentro de um conjunto, em um processo de clusterização, isto é, o dividir os dados em agrupamentos (*clusters*). Há, portanto, maior ingerência humana no primeiro processo.

Quanto às classes de tarefa que um algoritmo de *machine learning* pode executar, os exemplos mais comuns são: i) *regressão*: encontrar um valor contínuo (e.g. um programa que aprende a definir preços de imóveis a partir de registros

¹⁹ RUSSEL, S., op. cit., p. 7

²⁰ Tradução livre: “Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed” em PUGET, Jean François. What Is Machine Learning?. Disponível em: <https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en>. Acesso em: 19 nov. 2019.

²¹ PUGET, Jean François, op. cit.

²² PUGET, Jean François, op. cit.

imobiliários anteriores); ii) *classificação*: escolher uma resposta dentro de um grupo finito de opções (como prever se uma transação de cartão de crédito é ou não fraudulenta); iii) *segmentação*: de forma não supervisionada, classificar em *clusters* exemplos similares (agrupar consumidores com padrões de compras semelhantes); e *análise de redes*: bucar entrelaçamentos (nós) importantes dentro de redes (ferramentas de pesquisa, como o Google, utilizam esse tipo de programa para classificar redes formadas por páginas e seus hiperlinks).²³

Importa salientar que há várias abordagens de *machine learning*, sendo que uma das mais populares na atualidade é o aprendizado profundo (*deep learning*)²⁴, que está intimamente relacionada às redes neurais artificiais (*artificial neural networks* - ANN). O que caracteriza o aprendizado profundo é que o próprio algoritmo detecta seus erros e realiza os ajustes necessários para aprimorar seus resultados²⁵.

Sua grande vantagem, portanto, é não precisar da intervenção de um especialista para realizar tarefas de grande complexidade, ou melhor, justamente por não depender de um humano para orientá-lo na execução das atividades é que ele está apto a resolver problemas que nem mesmo os seres humanos são capazes de explicar com exatidão.

Assim, o *deep learning* é especialmente útil em problemas complexos (ao menos para máquinas), tais como classificação de imagens, processamento de linguagem natural e reconhecimento de fala.²⁶ Desenvolvê-lo, contudo, requer uma grande quantidade de dados e grande poder de processamento, sendo esse um dos motivos pelos quais só recentemente foi possível viabilizá-lo em grande escala.

Tais modelos, no mais das vezes, operam através de redes neurais artificiais, que consistem em unidades computacionais, chamadas de neurônios artificiais, conectadas entre si e divididas em camadas. Os dados de entrada (*inputs*) ativam

²³ PUGET, Jean François, op. cit.

²⁴ Outros modos de se obter aprendizado de máquina podem ser: árvores de decisão, programação lógica indutiva, clusterização, redes Bayesianas, aprendizagem por reforço, dentre outras.

²⁵ GROSSFELD, Brett. A simple way to understand machine learning vs deep learning. Zendesk. Disponível em: <<https://www.zendesk.com/blog/machine-learning-and-deep-learning/>>. Acesso em: 19 nov. 2019.

²⁶ MAHAPATRA, Sambit. Why Deep Learning over Traditional Machine Learning?. Disponível em: <<https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>>. Acesso em: 11 nov. 2019.

uma primeira camada, que ativa a camada seguinte, assim por diante, até chegar a um valor de saída. Tais redes artificiais são inspiradas, ainda que de modo rudimentar, em redes neurais biológicas, cujo funcionamento ainda não é claro para seres humanos.

Originalmente, devido a limitações de ordem prática, redes neurais possuíam poucas camadas (uma de entrada e outra de saída). Contudo, avanços tecnológicos permitiram que fossem introduzidas camadas intermediárias (*hidden layers*), que aumentam o desempenho de algoritmos na resolução de problemas complexos. Cada camada representa um nível na hierarquia e é responsável por uma tarefa específica (e.g. uma é responsável por encontrar pontos, a próxima identifica linhas, depois formas geométricas e a última uma imagem completa). Desse processo, com camadas “ocultas” e hierarquicamente divididas, origina-se o nome “aprendizado profundo” ou *deep learning*.

Foi por meio do *deep learning* que algoritmos foram capazes de realizar feitos notórios, tais como o *Deep Blue* e o *AlphaGo*, que derrotaram campeões mundiais humanos em jogos até então tidos como preponderantemente humanos (xadrez e Go, respectivamente)²⁷. Isso porque seu aprendizado hierárquico e dividido em camadas o permite fazer múltiplos ajustes para aprimorar seu desempenho, desde que sejam disponibilizados dados suficientes. No próximo tópico, serão abordados aspectos próprios dessa abordagem.²⁸

1.1.2 Como máquinas “aprendem”: peculiaridades do *machine learning*

Normalmente máquinas são programadas através de algoritmos que explicam passo a passo como elas devem desempenhar uma tarefa (se isso, então aquilo). Contudo, certos problemas ou são muito complexos ou as pessoas simplesmente não sabem como resolvê-los. Por exemplo, apesar da maior parte dos seres humanos não ter o menor problema para identificar números escritos à mão,

²⁷ Cf. ROSA, Natalie. Ex-campeão de Go disse ter se aposentado por ser impossível vencer a IA. 2019. Disponível em: <<https://canaltech.com.br/inteligencia-artificial/ex-campeao-de-go-disse-ter-se-aposentado-por-ser-impossivel-vencer-a-ia-156451/>>. Acesso em: 30 nov. 2019.

²⁸ MAHAPATRA, Sambit, op. cit.

ainda assim não se sabe ao certo como o cérebro processa a informação recebida através dos nervos ópticos e a interpreta de forma a reconhecer um certo padrão de traços como um três, por exemplo.

O mesmo ocorre com o reconhecimento de sons, imagens, expressões faciais, dentre outros. Além disso, jogos como o Go possuem mais resultados possíveis do que há átomos no universo, de forma que tentar calcular as todas as possibilidades de jogadas possíveis é inviável.²⁹ Se não se sabe explicar, também não é possível programar uma máquina para fazê-lo, ao menos não de forma direta. Entretanto, como dito, através do *machine learning* pode-se criar programas que buscam padrões e aprimoram seu desempenho através de testes, após múltiplas repetições, e que fazem ajustes de forma automática.

Em termos matemáticos, trata-se de encontrar e otimizar uma função, ou seja, algo que descreve numericamente a relação dos *inputs* com os *outputs*. Em essência, o que caracteriza todos os algoritmos de aprendizado de máquina é “otimizar um critério de performance utilizando dados ou experiências anteriores”³⁰.

O processo de “aprendizado”, em uma rede neural artificial, ocorre através da manipulação de pesos e vieses.³¹ Cada neurônio possui um viés (valor mínimo) de ativação e está conectado a todos os neurônios da próxima camada. Para ser ativado, e com isso propagar informação para o próximo “neurônio”, o impulso (*input*) recebido deve ser suficientemente forte (estar acima de um determinado valor) para romper tal limiar, de forma que ele gera um *output*, que servirá para ativar (ou não) o próximo neurônio da rede, em um processo em cascata.

Além dos vieses, cada uma dessas conexões possui “pesos”. Dessa forma um impulso mais fraco pode ser potencializado por uma conexão de maior força (e assim ativar o próximo neurônio), ao passo que um *output* mais forte pode ser

²⁹ GERMANO, Felipe. Computador vence humano em Go, jogo mais complexo que xadrez. Super Interessante. São Paulo. 04 nov. 2016. Disponível em: <<https://super.abril.com.br/tecnologia/computador-vence-humano-em-go-jogo-mais-complexo-que-xadrez/>>. Acesso em: 19 nov. 2019.

³⁰ COGLIANESE, Cary and LEHR, David. Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, pp. 1156 - 1160

³¹ Aqui a palavra “viés” é utilizada em sentido diferente do que será utilizada no resto deste trabalho, especialmente no Capítulo 3. Nesta seção, “viés” se refere a um valor numérico objetivo, enquanto, nas demais, quer dizer uma distorção ou resultado indesejado, geralmente consequência da “contaminação”, voluntária ou involuntária, de um algoritmo por opiniões ou posicionamentos pessoais da sociedade e/ou de seus programadores.

mitigado se sua conexão tiver um peso menor. “Aprender”, no caso, significa otimizar sua performance, *sem intervenção humana*, manipulando tais valores numéricos, o que servirá para fazer predições em novos dados.

A partir do exposto, é possível elencar algumas das principais características desse tipo de algoritmo. Em primeiro lugar, assume-se que os dados ainda não vistos pelo algoritmo são similares aos dados que os pesquisadores utilizaram para treiná-lo.³² Trata-se, portanto, de um raciocínio indutivo, o qual está sujeito a todas as limitações inerentes a esse método.³³

Em segundo lugar, nas abordagens estatísticas tradicionais, os pesquisadores precisam especificar uma equação matemática para expressar o resultado de uma variável como uma função de outras variáveis previamente escolhidas. Trata-se de estimar a magnitude e a direção das relações entre as variáveis escolhidas.³⁴ Em contraste, no aprendizado de máquina, não é necessário especificar um modelo prévio, o que acaba sendo justamente o que lhe permite desempenhar tarefas que os seres humanos não conseguem.

Ademais, a ausência de um padrão lógico (ao menos para um ser humano) o leva a ter um maior desempenho que outras técnicas não inteligentes, pois as IAs podem se adaptar mais facilmente e optar entre diferentes alternativas (tentativa e erro, julgamento técnico, estatística tradicional, etc.) para otimizar sua performance.

Todavia, o ganho em desempenho vem com contrapartidas. Uma delas é que o modelo será inútil para desempenhar qualquer outro trabalho, ainda que muito similar ao primeiro, pois sua configuração é demasiado específica, sendo necessário treiná-lo novamente quando da menor alteração na tarefa a ser desempenhada.

Outra perda ocorre no aspecto interpretativo. Os padrões que o algoritmo encontra não seguem necessariamente relações causais que ocorrem no mundo real (inferências causais). É daí que se diz que tais algoritmos são uma “caixa-preta” (*black-box*), pois “tudo o que se pode afirmar é que o algoritmo transforma uma série

³² PUGET, Jean François, op. cit.

³³ Por exemplo, pode-se falar do fenômeno do sobreajuste (*overfitting*), que consiste em criar relações absurdas, tais como prever o número de acidentes de carro na Lituânia a partir do tamanho da Floresta Amazônica e da população da Alemanha. Cf. DAHMS, Florian. What is "overfitting"? 2016. Disponível em: <<http://fdahms.com/2016/05/11/overfitting-explained/>>. Acesso em: 01 nov. 2018.

³⁴ COGLIANESE, op. cit., pp. 1156-1157.

de valores de entrada em valores de saída através da otimização de um critério de performance.”³⁵

Por conta disso, os usuários não conseguem precisar quais relações entre variáveis são levadas em conta pelo algoritmo, nem como ele classifica tais relações para criar classificações.³⁶ Ainda assim, não poder ver o que acontece dentro da caixa não significa que não haja ingerência humana nesse processo, tampouco que seus resultados não sejam minimamente interpretáveis. Ainda cabe aos pesquisadores escolher quais dados serão utilizados para treinar a máquina, estabelecer sua definição de sucesso e auditar seus resultados.

Feitos tais esclarecimentos, é necessário fazer a ressalva de que este trabalho utilizará, nos próximos capítulos, com o fim de tornar a leitura menos repetitiva, os termos “robôs” ou “máquinas” (e eventuais correlatos) como sinônimos de “algoritmos” ou “modelos”, a depender do contexto. Ademais, pelas mesmas razões, quando se utilizar “inteligência artificial”, na verdade, estar-se-á referindo a uma de suas abordagens, a saber, o aprendizado de máquina.

Por fim, salienta-se que, dentre as diversas possíveis aplicações de inteligência artificial, a que será mais estudada neste trabalho é a do processamento de linguagem natural (PLN ou NLP), que consiste nos esforços voltados a ensinar algoritmos a “interpretar” a linguagem humana, processo que, como será visto, apesar de conter algumas semelhanças, diferencia-se substancialmente da forma como humanos entendem línguas naturais.

1.2 Aplicando aprendizado de máquina a textos legais

1.2.1 Especificidades do campo jurídico

Há grande interesse e motivos para que se busque aplicar ferramentas de inteligência artificial no campo do direito. Todavia, ainda que algoritmos de NLP sejam capazes de “ler” textos de maneira relativamente satisfatória, espera-se mais que isso nas atividades jurídicas. Certos programas, por exemplo, são capazes de

³⁵ COGLIANESE, op. cit., p. 1158.

³⁶ COGLIANESE, op. cit., pp. 1159-1160.

escrever notícias relacionadas a esportes, ao passo que o *Watson* e o *Debater*, ambos da IBM, são capazes, respectivamente, de responder perguntas pontuais e extrair argumentos de textos. Ainda assim, não seria possível utilizá-los para solucionar problemas legais, uma vez que tais programas não são capazes de performar o raciocínio legal (*legal reasoning*).³⁷

Para que um programa de computador seja capaz de desempenhar atividades jurídicas ele deve possuir meios de não apenas responder questões, mas também de explicar suas respostas de forma inteligível aos profissionais do direito.³⁸ Ademais, para se formular argumentos legais, é necessário seguir certos padrões de estruturação, identificação e relação entre conceitos, como, por exemplo, padrões probatórios e a autoridade de normas legais ou decisões de tribunais superiores.³⁹

Ainda assim, o *Watson* e o *Debater* exemplificam como programas de computador podem vir a auxiliar humanos a formular e testar hipóteses legais, ao passo que lançam as bases de algumas tecnologias relativas à extração de informações de textos, que podem ser utilizadas para alimentar modelos computacionais. Como resultado, inaugura-se a computação cognitiva (*cognitive computing*) ou a corrobótica (*co-robotics*), a saber, um trabalho colaborativo entre humanos e máquinas, através do qual cada um realiza o tipo de atividade que melhor é capaz de desempenhar.

Um modelo computacional de performance de raciocínio jurídico (*computational model of legal reasoning* - CMLR) consiste em um programa de computador que “quebra” um processo intelectualmente complexo em um conjunto de instruções executáveis por um programa de computador, evidenciando os elementos humanos do *legal reasoning*. Anteriormente à disseminação de técnicas de aprendizado de máquina, era necessário “extrair” manualmente das fontes de direito as informações que alimentam tais modelos e traduzí-las em linguagem de programação, o que limitava substancialmente a viabilidade prática desses processos.⁴⁰

³⁷ ASHLEY, Kevin D., op. cit., p.14.

³⁸ ASHLEY, Kevin D., op. cit., p. 23.

³⁹ ASHLEY, Kevin D., op. cit., p. 26.

⁴⁰ ASHLEY, Kevin D., op. cit., p. 3-4.

Com o implemento de tecnologias de *question-answering* (sistema de perguntas-respostas - QA), *information extraction* (extração de informações - IE) e *argument mining* (mineração de argumentos), que em conjunto formam as bases da *mineração de textos (text mining)*, o processo de extração de informações pode ocorrer de forma automatizada. Um sistema QA tem por objetivo encontrar, dentro de um grande texto ou conjunto de textos, uma frase que responde precisamente à pergunta de um usuário. A extração de informações consiste em resumir detalhes essenciais específicos a um documento em particular. A mineração de argumentos diz respeito à identificação de estruturas argumentativas em um texto, tais como premissas, conclusões e relações entre argumentos. Aplicada a documentos legais, a mineração de textos consiste na *análise de dados textuais legais (legal text analytics)*.⁴¹

Outras duas ferramentas também são de grande relevância para viabilizar a resposta de perguntas jurídicas, a saber, a expansão de conceitos (*concept expansion*) e extração de relações (*relation extraction*). Ocorre que certos textos jurídicos (e.g cláusulas em um contrato) podem ser expressos de formas bastante diferentes como ocorre em “a empresa X não se responsabiliza por eventuais despesas decorrentes...” e “quaisquer custos serão arcados integralmente pela empresa Y...”. Para se identificar que tais sentenças dizem, no fundo, a mesma coisa, é necessário expandir o significado de um termo, abrangendo outros que estão a ele relacionados, como ocorre com *custo, despesa, gasto, expensas* (expansão de conceitos), bem como identificar relações semânticas, ainda que as estruturas gramaticais da oração variem: “X não vai pagar Z” ou “Z será pago por Y” (extração de relações).⁴²

Normalmente, todas essas técnicas utilizam, em uma ou mais etapas, aprendizado de máquina para identificar relações semânticas entre os termos de um texto. Através delas é possível encontrar informações de forma mais eficiente e extrair automaticamente os dados utilizados na implementação de modelos computacionais de argumentação legal, auxiliando humanos em tarefas legais, tais como formular argumentos ou prever decisões.

⁴¹ ASHLEY, Kevin D., op. cit., p. 5.

⁴² ASHLEY, Kevin D., op. cit., p.28-29.

Isso se torna possível uma vez que o aprendizado de máquina extrai padrões de um conjunto de documentos relativos a um domínio específico, o que permite processar textos para identificar elementos relevantes à resolução de um certo problema e classificar as possíveis soluções por ordem de relevância, de forma automática.⁴³ Para se compreender esse processo de forma mais detalhada, é necessário analisar a forma como os textos são representados, quando da aplicação do ML.

1.2.2 Vetorizando textos

Algoritmos estabelecem relações entre *features* (aspectos dos dados) e um resultado (*outcome*). Possíveis *features* de um texto são: tipo, quantidade e posição de termos empregados, tipos de construções gramaticais contidas, informações semânticas e sintáticas e assim por diante. Possíveis resultados consistem no tipo de decisão (i.e. favorável, desfavorável), tipo de documento (petição inicial, contestação, recursos, etc.), relevância de um documento para um determinado fim (um *score*, ou uma classificação binária do tipo “relevante/não relevante”). Um dos principais objetivos da computação cognitiva é identificar *features* textuais relevantes para auxiliar humanos na solução de problemas.⁴⁴

Em breve síntese, para se proceder à análise artificial de textos jurídicos, é necessário transformar um *corpus textual* (um conjunto de dados textuais, tais como o acervo de processos de um tribunal), em um espaço vetorial multidimensional, no qual um algoritmo de ML possa trabalhar. O primeiro passo consiste em coletar e processar dados não tratados (ou “crus”), ou seja, textos jurídicos em linguagem natural. Em seguida, é necessário “tratar” ou “normalizar” tais dados através de diferentes processos. Isso é feito transformando-se todas as letras em minúsculas (assim, “Direito” torna-se “direito”) e reduzindo as palavras às suas raízes inflexionadas (“pagaram” e “pagou” tornam-se “pag”). A *tokenização* remove caracteres especiais do texto, tais como acentos, hífens, pontuação, bem como

⁴³ ASHLEY, Kevin D., op. cit., p. 13.

⁴⁴ ASHLEY, Kevin D., op. cit., p. 235-236.

certas palavras que são muito repetitivas e possuem pouco ou nenhum significado, tais como “e” “do”, “a”, “quem”, assim chamadas as “*stop words*”⁴⁵.

Esse processo também diz respeito ao tratamento de termos vizinhos⁴⁶, os quais podem ser tratados aglutinados em tokens de n termos (n-gramas), isto é, a frase “carro trafegando em alta velocidade” pode ser representada em trigramas como “carro trafegando em”, “trafegando em alta”, “em alta velocidade”. A *anotação* auxilia na desambiguação de termos similares, inserindo informações úteis para definir o sentido em que uma palavra foi utilizada no texto, tal como sua classificação morfológica (verbos, substantivos, adjetivos ou advérbios) ou sintática (sujeito, objeto direto, objeto indireto, etc).⁴⁷

Após tais processos, um documento é transformado em um vetor de aspectos (*feature vector*). Tais vetores contêm características daquele determinado texto, tais como seus n-gramas, quantidade de palavras e informações relativas à sua categoria. O vetores de todos os textos possuem todos o mesmo comprimento n, correspondente ao número total de palavras e outras *features* do *corpus textual*.

Assim, cada documento é representado como uma entidade num espaço de n dimensões, pois cada um de seus aspectos servirá para definir sua magnitude em cada uma das direções do espaço vetorial, de forma que se pode diferenciar um documento de outro através do “local” em que ele se encontra. Tome-se como exemplo um *corpus* de apenas 3 documentos, cada um podendo conter até 3 aspectos (*features*), distribuídas conforme a seguinte tabela:

	<i>Feature 1 (x)</i>	<i>Feature 2 (y)</i>	<i>Feature 3 (z)</i>
Documento A	Sim	Não	Sim
Documento B	Não	Sim	Sim
Documento C	Sim	Sim	Não

⁴⁵ Não há critérios únicos para a tokenização, certos programas podem adotar diferentes critérios para remover caracteres especiais e *stop words*, visando manter certas distinções semânticas. Tais regras devem ser analisadas na aplicação do ML a textos jurídicos, uma vez que podem remover informações relevantes (com conteúdo significativo). (p. 236)

⁴⁶ O mesmo pode ser feito, como se vê no item 3.1, com palavras isoladas, que podem ser classificadas em n-gramas de diversos tamanhos.

⁴⁷ ASHLEY, Kevin D., op. cit., p. 236-237.

Sua representação, em um espaço vetorial de tridimensional, dá-se da seguinte forma:

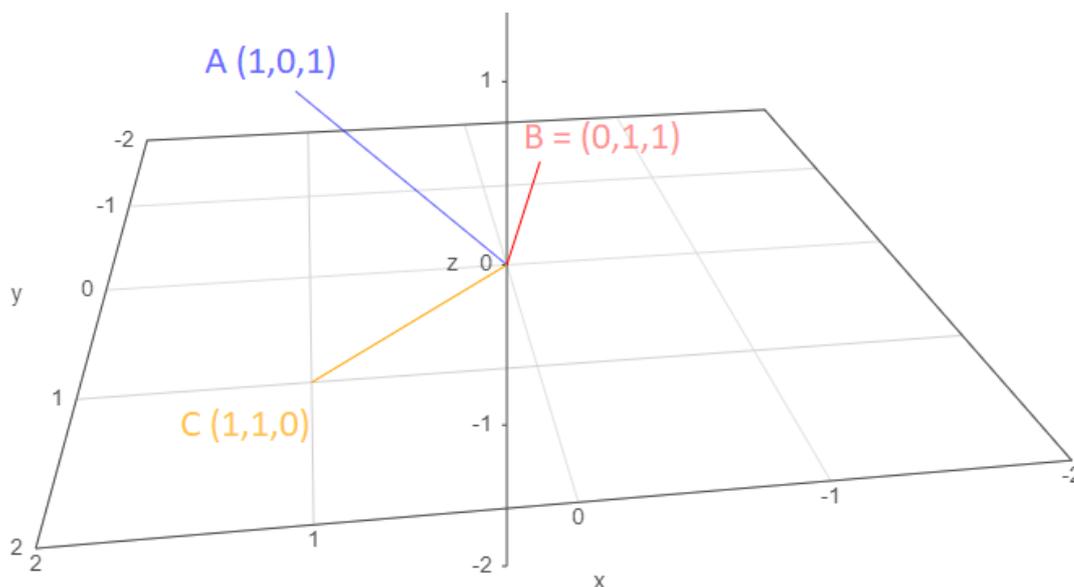


Figura 1 - Representação vetorial de três documentos a partir de suas *features*.

Assim, percebe-se que documentos que possuem aspectos parecidos tenderão a ocupar a mesma região do espaço vetorial. Deve-se observar que o número n de dimensões normalmente será grande, pois conterà, no mínimo, o número de diferentes palavras ou n -gramas existentes em todo o *corpus*. A magnitude de cada aspecto é variada, podendo ser simplesmente “0” ou “1”, à medida que o documento contém ou não aquele aspecto, ou indicar quantidade de vezes que a palavra ou aspecto se repete no texto, bem como seu valor *tf/idf* (*term frequency–inverse document frequency*)⁴⁸, a saber, a frequência com que a *feature* aparece no documento descontada da sua frequência no *corpus*.⁴⁹

Em seguida, dividem-se os documentos em dados de treino e de teste. Pode-se fazer isso selecionando aleatoriamente alguns para compor os dados de teste e utilizando os demais para “treinar” o algoritmo. Em um processo de validação

⁴⁸ A ideia desse indicador é medir o “peso” de uma palavra para um documento comparado com seu peso no *corpus*. Assim, seu valor aumenta proporcionalmente à frequência dela naquele documento e diminui à medida que outros documentos a contém. Se uma palavra aparece muitas vezes no documento em análise, mas ela também aparece em muitos outros, é um sinal de que ela não tem muita relevância para sua classificação, logo seu *tf/idf* é baixo.

⁴⁹ ASHLEY, Kevin D., op. cit., p. 237.

cruzada, utilizado quando a quantidade de dados é relativamente pequena, eles são divididos em subconjuntos, que serão utilizados ora como dados de treino, ora como de teste. Por fim, a avaliação dos resultados obtidos se dá, dentre outras formas, através da medição de sua acurácia, ou seja, a razão entre os resultados corretamente classificados e o número total de instâncias de teste.⁵⁰

Tais algoritmos têm por fim realizar uma classificação ou predição. Como exemplos de classificação, pode-se distinguir entre documentos relevantes e irrelevantes para um determinado caso, diferenciar trechos de decisões judiciais (e.g. relatório, fundamentação e dispositivo de um sentença), identificar estruturas argumentativas em documentos legais, bem como indicar a pertinência de um processo a um determinado tema repercussão geral ou recurso repetitivo. Com relação à predição, poderia-se utilizar o aprendizado de máquina para indicar a probabilidade de êxito de uma determinada ação, comparando processos similares ou levando em conta o histórico de decisões de um determinado juiz ou tribunal.

A importância do aprendizado de máquina nesse processo consiste em estimar o peso relativo de cada um das centenas ou milhares de *features* que um documento possui. Ainda que os humanos tenham uma noção intuitiva de quais aspectos possuem relevância, não é possível mensurar quais terão maior importância, nem em que grau isso ocorre.

Assim, utiliza-se o ML para calcular quais relações entre as variáveis mais se aproximam de explicar os dados de treino e de teste. É nisto, portanto, que reside sua autonomia: jogar com as *features*, através da tentativa e erro, para otimizar a função que melhor descreve os dados que o algoritmo possui à sua disposição.

Uma das formas de se chegar a tais classificações/predições é a partir de uma *máquina de vetores de suporte* (*support vector machine - SVM*). Ao contrário de outros classificadores como, por exemplo, *árvores de decisão*, as SVMs utilizam modelos estatísticos complexos e seus resultados não são facilmente interpretáveis por seres humanos.

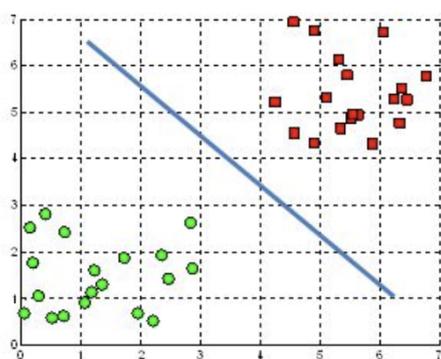
Como dito, quando se aplica o aprendizado de máquina a textos jurídicos, os documentos são transformados em entidades em um espaço vetorial. Se os aspectos de diferenciação utilizados forem minimamente efetivos, entidades

⁵⁰ ASHLEY, Kevin D., op. cit., p. 238.

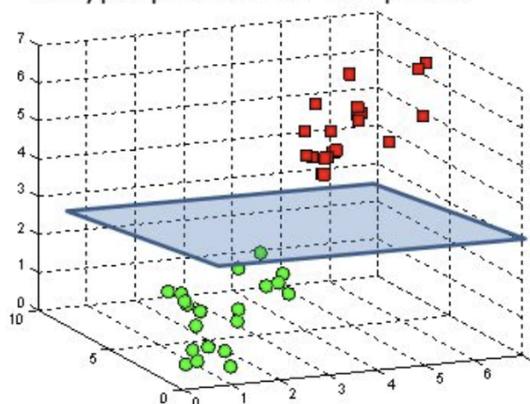
similares tenderão a estar posicionadas próximas umas das outras nesse espaço n -dimensional.

Assim, o objetivo de uma SVM é, através de critérios estatísticos, criar uma fronteira que separe instâncias positivas de negativas de uma certa classe (e.g. relevante e não relevante). Essa fronteira será uma entidade geométrica, chamada de hiperplano, e sua natureza dependerá da quantidade de dimensões do espaço vetorial. Em um espaço unidimensional, a fronteira será um ponto, já em um plano, a fronteira será uma linha e, em um espaço tridimensional, ela será um plano. Para fins didáticos, normalmente se ilustram espaços vetoriais de até três dimensões, todavia, em dados reais, cada *feature* será uma dimensão, o que significa espaços de centenas ou milhares de dimensões. Isso mostra a dificuldade para que os resultados se tornem inteligíveis a seres humanos.

A hyperplane in \mathbb{R}^2 is a line



A hyperplane in \mathbb{R}^3 is a plane

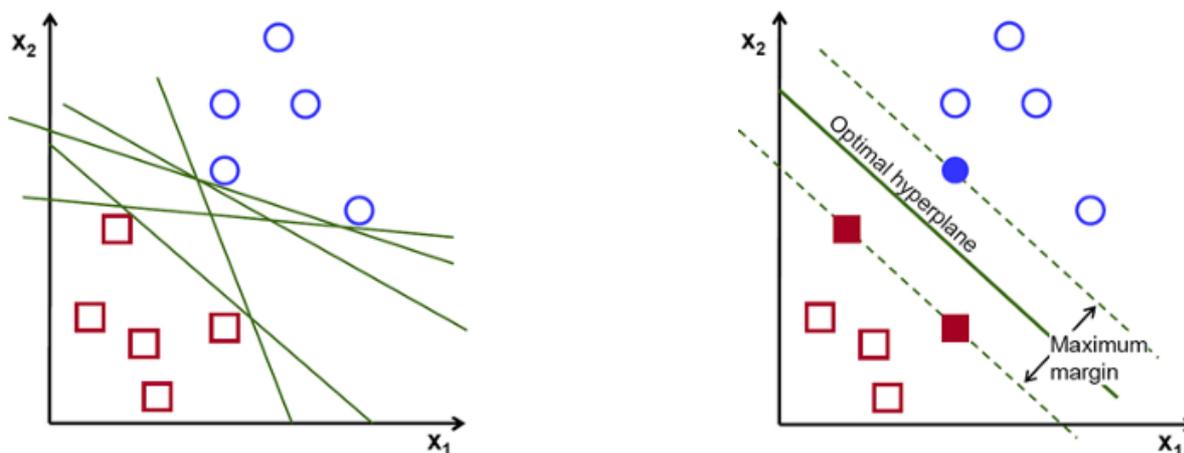


Figuras 2 e 3 - Ilustrações de hiperplanos em espaços, respectivamente, bi e tridimensionais.⁵¹

Uma vez calculado o hiperplano, classificar dados futuros, ou seja, que não foram utilizados como treino ou teste, será simplesmente uma questão de verificar de qual lado da fronteira a nova entidade se posiciona. Dentre as diferentes

⁵¹ GANDHI, Rohith. Support Vector Machine — Introduction to Machine Learning Algorithms. Disponível em: <<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>>. Acesso em: 16 nov. 2019.

maneiras de traçar essa fronteira, “a SVM escolhe a que maximiza a margem entre a fronteira e as demais instâncias”. Ocorre, todavia, que classes de dados reais frequentemente não podem ser perfeitamente separadas de um lado ou de outro da fronteira, de forma que a SVM precisa calcular uma “margem de tolerância” para casos erroneamente classificados.⁵²



Figuras 4 e 5 - Diferentes vetores de suporte possíveis para separar instâncias positivas e negativas de uma classe e hiperplano ótimo.⁵³

Dessa forma, pode-se ter uma ideia geral de alguns conceitos básicos do campo da inteligência artificial e de como é possível aplicar algoritmos de aprendizado de máquina a textos jurídicos, sendo que foge do escopo deste trabalho explicar esse processo em maiores detalhes. Todavia, para se ter uma melhor compreensão das consequências do uso desse tipo de ferramenta na prática jurídica, deve-se analisar a questão dos “juízes-robôs” a partir de outros pontos de vista, que não o estritamente técnico.

⁵² ASHLEY, Kevin D., op. cit., p. 251.

⁵³ GANDHI, Rohith. Support Vector Machine — Introduction to Machine Learning Algorithms. Disponível em: <<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>>. Acesso em: 16 nov. 2019.

Capítulo 2 - A linguagem no direito: pragmática e discricionariedade

A o uso do aprendizado de máquina no direito envolve também compreender os limites da própria linguagem natural, ou seja, aquela falada cotidianamente por seres humanos, através da qual leis são estatuídas, argumentos, formulados e decisões, escritas. Isso porque a forma como se entende a linguagem influencia na própria maneira de se conceber o fenômeno jurídico.

Defende-se aqui que tanto a atividade do juiz, quanto dos robôs, estarão sujeitas à imprecisão da linguagem natural, abrindo espaço para casos em que há margens de interpretação. Por conta disso, ainda que as diferentes teorias jurídicas aqui abordadas dêem diferentes contornos e definições à atividade jurídica discricionária, nenhuma delas será capaz de fugir da problemática da pluralidade de decisões válidas para um mesmo caso. Ainda assim, para melhor se entender tais discussões, é necessário partir de seu pano de fundo, que se dá com a filosofia da linguagem.

A partir da filosofia analítica do século XX, operou-se a chamada “virada linguística”, por meio da qual “a filosofia só poderia ser pensada a partir da linguagem”⁵⁴, movimento que também ecoou na teoria do direito. Os problemas filosóficos afastaram-se dos sistemas metafísicos e passaram a ser descritos em torno da crítica da linguagem, ao mesmo tempo que os juristas passaram a tentar responder perguntas da filosofia do direito através de conceitos linguísticos.

Nesse contexto, adquirem especial relevância as obras de Ludwig Wittgenstein, cuja fase inicial do pensamento ilustra uma visão mais “logicista” da linguagem, que serviu de base para o positivismo científico, também chamado de neopositivismo ou empirismo lógico. Sua obra tardia, entretanto, indo na direção oposta, privilegia as estruturas e funções da linguagem comum, tomando como objeto privilegiado de sua análise o contexto em que um discurso é proferido, ou seja, seu âmbito pragmático.

⁵⁴ SIMON, Henrique Smidt. Direito, Hermenêutica e Filosofia da Linguagem, p. 15.

Foi Charles Morris o responsável pela divisão da semiótica em três dimensões: sintaxe, semântica e pragmática.⁵⁵ Como visto na introdução, a semióse consiste “no processo em que algo funciona como um signo”, movimento que possui três fases: veículo sígnico, *designatum* (aquilo a que o signo se refere) e intérprete. Das diferentes relações entre esses elementos, podem-se extrair as diferentes dimensões do processo semiótico.

A sintaxe diz respeito à relação dos signos entre si. É nesse âmbito que são estudadas as regras de formação e transformação de proposições, através das quais organizam-se os elementos de uma língua e torna-se possível inferir sentenças a partir de outras sentenças. Como será visto, foi à sintaxe de linguagens científicas que o Círculo de Viena dedicou a maior parte de seus esforços.

A semântica, ao seu turno, se preocupa com os significados, o que remete à relação entre os veículos sígnicos e os objetos aos quais eles se referem, seus designata. É por meio de uma regra semântica que se definem as condições em que um signo se aplica a objetos ou situações. A terceira dimensão, pragmática, compreende a análise contextual, ou seja, a forma como os signos são empregados por aqueles que os utilizam. No processo semiótico, os significados são circunstanciais, pois os signos adquirem sentido por meio de seu uso em diferentes situações.⁵⁶

A filosofia da linguagem influenciou a forma como Hans Kelsen e Herbert Hart pensaram o fenômeno jurídico, utilizando conceitos que se aproximam, respectivamente, da primeira e da segunda fase da obra de Wittgenstein.⁵⁷ Ainda assim, como será explorado ao longo deste capítulo, dados os pressupostos da filosofia analítica da linguagem, nos quais ambos os teóricos do direito estão inseridos, a questão do decisionismo, aqui entendido como uma espécie de discricionariedade que o operador do direito possuirá quando da aplicação da norma ao caso concreto, mostra-se incontornável.⁵⁸

Se corretas tais ponderações, elas impactam diretamente na forma de se entender algoritmos de processamento de linguagem natural, especialmente no

⁵⁵ FIDALGO, António. *Semiótica: A Lógica da Comunicação*, p. 93.

⁵⁶ FIDALGO, António. *Semiótica: A Lógica da Comunicação*, pp. 95-102.

⁵⁷ SIMON, Henrique Smidt, *op. cit.*, p. 15.

⁵⁸ SIMON, Henrique Smidt, *op. cit.*, p. 16.

âmbito jurídico. Ora, se a linguagem humana é imprecisa e, por conta disso, nem sempre é possível definir com precisão o significado de um termo, em alguns casos será necessário arbitrar uma definição. Sob essa perspectiva, robôs têm sua suposta imparcialidade questionada, vez que eles estarão apenas importando dos seres humanos certas definições linguísticas ou legais.

Nas próximas seções, será abordada a questão do decisionismo tanto na obra de Kelsen e na de Hart, para, em um próximo momento, explicar-se em linhas gerais como algoritmos de aprendizado de máquina são capazes de captar o contexto, buscando-se responder a questão de saber se eles estariam “aptos”, ainda que minimamente, a assimilar diferentes modos de uso dos termos. Tais reflexões tornam-se necessárias para melhor explorar o conteúdo do último capítulo deste trabalho, que colocará em foco como heurísticas e vieses afetam humanos e também máquinas.

2.1 O positivismo científico e a Teoria Pura do Direito de Kelsen: a visão descritivista da linguagem

As ideias do jurista austríaco Hans Kelsen estão inseridas nos pressupostos do neopositivismo, cujos fundadores reconhecidos são os integrantes do chamado Círculo de Viena. Tal corrente linguístico-epistemológica afirma que a linguagem não apenas permite a troca de informações, mas também funciona como meio de controle dos conhecimentos humanos. Seu membros mais destacados foram, dentre outros, Nagel, Moris, Quine e Fèigl e figuram entre seus precursores o Wittgenstein inicial, Frege e Russell.⁵⁹

Gottlob Frege, pai da lógica moderna, é considerado o responsável pela chamada virada linguística na filosofia, a saber, a busca sistematizada de esclarecer problemas filosóficos através do tratamento de questões linguísticas. Trata-se de um projeto que busca uma linguagem logicamente precisa, para se evitar as ambiguidades da linguagem natural, de forma a possibilitar maior controle sobre o desenvolvimento da ciência e da matemática.

⁵⁹ Cf. WARAT, Luis Alberto. O Direito e sua linguagem, p. 37 e SIMON, Henrique Smidt, op. Cit., p. 21.

Dentre outras contribuições, tornou-se célebre sua distinção entre sentido e referência, que diferencia a forma como um objeto é apresentado (seu sentido) do objeto designado (sua referência, aquilo sobre o que se fala). Nessa perspectiva, para que um enunciado possa ter um valor de verdade, e, por consequência, transmitir conhecimento, ele deve afirmar algo sobre o mundo, ou seja, ter uma referência. Verdade, por consequente, diz respeito a uma correspondência (relação) entre “uma imagem e seu objeto”.⁶⁰

Ao lado de Frege e de Russell, a primeira fase do pensamento de Wittgenstein configura um dos pressupostos do neopositivismo. O *Tractatus logico-philosophicus*, primeiro e único livro publicado em vida pelo filósofo austríaco, pode ser interpretado como uma tentativa de elucidação de erros advindos da não compreensão das estruturas lógicas da linguagem comum, por conta da qual derivam os problemas da filosofia tradicional.

Nessa perspectiva, que assume uma forte postura antimetafísica, o mundo se apresenta de forma objetiva, à medida que corresponde à totalidade de estados de coisas, que, por sua vez, depende das combinações entre objetos independentes. Tudo que resta aos sujeitos é apreender a forma lógica de tais combinações, através da qual é possível que a linguagem reflita (figure) o mundo. A linguagem, portanto, é a forma de representação do mundo.

Dessa forma, Wittgenstein cria “uma ontologia que prescindir de qualquer característica subjetiva”⁶¹, uma vez que os objetos que compõem o mundo possuem tão somente propriedades formais e qualquer tentativa de falar sobre algo que não seja uma proposição empírica ou de um enunciado lógico (tautologia ou contradição) ultrapassará os limites da própria linguagem. Isso porque as proposições que desrespeitam tais limites nunca poderão ser nem verdadeiras nem falsas, e nada pode ser dito a esse respeito. Isso é justamente o que ocorre com enunciados metafísicos (como a ética, por exemplo, uma vez que esta depende da vontade individual), chamados de *pseudoproposições*, os quais compõem o “místico”, a parte inefável da linguagem (e do mundo).⁶²

⁶⁰ SIMON, Henrique Smidt, op. cit., p. 31 e 32.

⁶¹ SIMON, Henrique Smidt, op. cit., p. 25.

⁶² SIMON, Henrique Smidt, op. cit., p. 26 e 27.

Ainda assim, demonstrar essa forma lógica ou a semântica não é possível, pois seria necessário, para tanto, sair do mundo e, conseqüentemente, sair da lógica, porém ocorre que não há possibilidade de pensamento ou linguagem fora desta.⁶³ A forma lógica, assim, simplesmente se mostra através de uma determinada estrutura, não podendo ser dita ou demonstrada. Dessa forma, como não cabe ao filósofo teorizar sobre o místico, o que inclui a própria forma lógica, resta à filosofia “apenas” a tarefa de elucidar as regras da linguagem ordinária, ou seja, eliminar falsos problemas, enquanto afirmações (empíricas) sobre o mundo estão restritas às atividades científicas⁶⁴.

Isso posto, torna-se fácil compreender as bases teóricas do positivismo lógico, que teve por fundador o Círculo de Viena. Tal corrente é caracterizada pela forte postura antimetafísica e pelas ponderações acerca da possibilidade do conhecimento e sua enunciação. Igualmente, entendem os membros do Círculo que a filosofia possui por objeto somente os problemas relativos à lógica da linguagem, pois está fora de suas atribuições formular sentenças com conteúdo empírico.

O conhecimento, dessa forma, só é possível a partir daquilo que é empiricamente apreensível e verificável. Sob essa ótica, a ciência adquire uma função privilegiada na formação do conhecimento, pois suas proposições descrevem fatos e suas hipóteses são controláveis ou verificáveis por meio da observação.⁶⁵

Consiste também em uma atitude típica dos “positivistas lógicos” a tentativa de solucionar problemas referentes aos limites dos enunciados através da criação de hierarquias de linguagem. Para tanto, utiliza-se a distinção entre *linguagem-objeto*, ou seja, a “linguagem em que se fala” e a *metalinguagem*, por meio da qual “se fala da linguagem objeto”. Essa seria, na perspectiva do *Tractatus*, uma alternativa ao “apelo a misteriosismos”, dos quais o próprio Wittgenstein inicial se valeu (e confessou)⁶⁶ para explicar os limites de sua própria obra, uma vez que suas próprias enunciações podem ser vistas como sem sentidos, já que afirmam

⁶³ Cf. WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus*, 3.03: “Não podemos pensar nada ilógico, porquanto, do contrário, deveríamos pensar illogicamente.”; 4.12: “A proposição pode representar a realidade inteira, não pode, porém, representar o que ela deve ter em comum com a realidade para poder representá-la — a forma lógica. Para podermos representar a forma lógica, deveríamos poder-nos instalar, com a proposição, fora da lógica, quer dizer, fora do mundo”.

⁶⁴ SIMON, Henrique Smidt, op. cit., p. 25.

⁶⁵ SIMON, Henrique Smidt, op. cit., p. 35-37.

⁶⁶ Cf. WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus*, 6.54 e CABRERA, op. cit., pp. 67-68.

coisas acerca da linguagem e não do mundo. Já para as ciências, a distinção entre metalinguagem e linguagem-objeto permite o estabelecimento de regras de controle dos processos elucidação e transformação controle da linguagem científica.⁶⁷

É esse o pano de fundo que influencia a *Teoria Pura do Direito* de Hans Kelsen. O Direito é “cientificado”, passando a ser regido pelo critério da verificabilidade e, para que isso se torne possível, cria-se uma distinção linguística, que difere os planos das normas, objetos por excelência da ciência normativa, e dos enunciados que as descrevem. Uma vez que as ciências devem descrever fatos empíricos, cabe ao cientista do direito apenas afirmar coisas acerca da validade de uma norma, uma vez que qualquer afirmação com relação ao seu conteúdo diz respeito a critérios subjetivos (metafísicos, portanto) ou a um critério de justiça universal, que, segundo o jurista austríaco, não existe, uma vez que a moral difere no tempo e no espaço.⁶⁸

Trata-se de uma busca por uma “univocidade lógico-conceitual”, que elimine da ciência jurídica quaisquer “compromissos políticos e efeitos ideológicos”⁶⁹, de maneira que o direito passa a ser entendido como a *forma jurídica válida*, independentemente do seu conteúdo, pois tal forma seria o único atributo comum às diversas estruturas normativas históricas, que permanece desde “uma tribo sob a liderança de um chefe despótico” até na “constituição da República suíça”. Isso configura a “auto-esgotabilidade do direito, o que, ao lado, da teoria piramidal, forma as bases da Teoria Pura do Direito.”⁷⁰

Uma norma, à medida que apenas prescreve uma conduta ao atribuir uma sanção à conduta oposta, não pode ser verdadeira ou falsa e não se aplicam a ela, conseqüentemente, relações lógicas. Todavia, para que um enunciado seja científico, ele deve ser capaz de gerar algum tipo de conhecimento, o que, para os neopositivistas, significa possuir um valor de verdade. Assim, para conferir ao direito o *status* de ciência, formulam-se proposições, às quais pode ser aplicada a lógica

⁶⁷ Cf. WARAT, op. cit., p. 48 e CABRERA, op. cit., p. 68.

⁶⁸ SIMON, Henrique Smidt, op. cit., pp. 86-87.

⁶⁹ WARAT, op. cit., p. 49.

⁷⁰ Cf. GONTIJO, Lucas de Alvarenga. Filosofia do Direito: metodologia jurídica, teoria da argumentação e guinada linguístico-pragmática, p. 23.

formal, ao passo que elas descrevem a existência e relações de uma norma (um ente ideal) e podem, por consequência, ser verdadeiras ou falsas.

Sua verdade ou falsidade, contudo, não diz respeito a um fato do *ser*, mas do *dever ser*. Proposições das *ciências naturais* estabelecem relações de *causalidade* entre seus objetos (fatos da realidade natural, tais como a propriedade de um metal se expandir ao ser aquecido), enquanto proposições jurídicas, que configuram uma *ciência normativa*, descrevem relações de *imputação*, ou seja, determinam a um fato da vida uma consequência, que pode ou não ocorrer.⁷¹

Como dito, para que uma norma seja válida e, portanto, exista, ela deve preencher o requisito da forma jurídica válida, o que significa, essencialmente, que ela foi elaborada conforme prescreve uma norma imediatamente superior na hierarquia normativa. Ao chegar-se ao topo dessa cadeia, que é ocupado pela primeira constituição histórica, Kelsen enuncia a *norma fundamental* como pressuposto de validade geral do ordenamento.⁷²

Interessante notar que, conforme elucida Warat⁷³, pode-se interpretar a teoria pura de Kelsen em termos da aplicação no campo jurídico dos conceitos de linguagem-objeto e metalinguagem. Utilizando-se a terminologia kelseniana, as *regras de direito* (enunciados jurídicos) são capazes de formular proposições descritivas, pois consistem na metalinguagem que tem por linguagem-objeto o *direito positivo*.

Isso tem importantes consequências para a noção de validade de uma norma, pois será através de um critério de decidibilidade metalinguístico, ou seja, uma regra que permite estabelecer se um enunciado faz parte ou não da linguagem-objeto, que será possível distinguir as normas que podem compor o direito positivo das demais. Tal critério nada mais é que a norma fundamental, cuja validade, nesses termos formulada, não pode ser questionada, pois não faz sentido questionar a validade de um critério de validez.

Isso porque, “quando se fala de um fundamento de validade, se está fazendo referência a um fundamento gnoseológico, a uma proposta significativa de metalinguagem e não a um processo de produção normativa”. O direito positivado,

⁷¹ KELSEN, Hans. Teoria Pura do Direito, pp. 53-54.

⁷² SIMON, Henrique Smidt, op. cit., 91, 93 e 95-97.

⁷³ WARAT, op. cit., pp. 48-52.

por sua vez, é a metalinguagem, através da qual se descrevem as condutas humanas (linguagem-objeto), dotando-lhes de significado objetivo, a saber, seu sentido social predominante, que é definido por sua significação normativa (deôntica).

2.2 Jogos de linguagem e o Conceito de Direito de Hart: a virada pragmática

Até então, nenhuma das análises linguísticas aqui expostas trouxe como objeto privilegiado de reflexão o campo da pragmática, limitando suas ênfases aos campos da semântica e da sintaxe. Isso ocorre porque o neopositivismo pressupõe que a linguagem comum é deficiente e seus erros devem ser suprimidos através da criação de uma linguagem formal.

Para a chamada “filosofia da linguagem ordinária”, cujo principal expoente foi Wittgenstein, em sua segunda fase, a filosofia deve esclarecer o uso da linguagem comum, adotando, para tanto, o âmbito pragmático da linguagem como nível de análise por excelência.⁷⁴ Privilegiar tal esfera da linguagem significa pressupor que “os significados socialmente padronizados possuem sentidos incompletos; são expressões em aberto, que apenas se tornam relativamente plenas em um contexto determinado”⁷⁵. Dessa forma, representa a virada pragmática uma expansão do conceito de linguagem, que passa a assimilar elementos decorrentes de interações humanas, reconhecendo a impossibilidade de uma linguagem logicamente perfeita e livre de vaguezas e ambiguidades.⁷⁶

Ainda dentro do paradigma da filosofia analítica, a segunda fase da obra de Wittgenstein, a saber, a das *Investigações Filosóficas*, altera consideravelmente a noção de linguagem do filósofo austríaco. Ainda assim, entende-se que as duas fases de seu pensamento mantêm alguns pontos em comum, tais como a recusa à metafísica, a função da filosofia como análise e esclarecimento da linguagem e as preocupações acerca do sentido linguístico.⁷⁷ A grande mudança reside, dessa forma, no que se refere à “lógica da linguagem”. Ao contrário do que fazem as

⁷⁴ WARAT, op. cit., pp. 62-64.

⁷⁵ WARAT, op. cit., pp. 65.

⁷⁶ SIMON, Henrique Smidt, op. cit., pp. 48-49.

⁷⁷ SIMON, Henrique Smidt, op. cit., p. 49.

visões filosóficas “ontologizantes”, que buscam encontrar a estrutura oculta que determina toda a linguagem, defende agora Wittgenstein que a tarefa da filosofia é descrever os diferentes usos da linguagem comum, que são plenamente acessíveis (não ocultos), sem prescrever uma teoria única, à qual todas as manifestações linguísticas devem se submeter.

Todavia, ainda que as proposições da linguagem comum carreguem imprecisões, isso não significa que não haja limites à utilização dos termos. Ocorre, na verdade, que não é possível estabelecer de antemão regras capazes de definir todos os casos em que um termo pode vir a ser empregado, de onde advém a importância da relação entre os signos e os sujeitos da comunicação, ou seja, a pragmática. Assim, somente o uso de uma expressão, em um dado contexto, é que será capaz de determinar seu significado.

Por conta disso, torna-se relevante o conceito de *jogo de linguagem*. Abandona-se a busca pela estrutura comum da linguagem e estabelece-se a ideia de *semelhança de família*: apenas é possível distinguir certos fenômenos que, por conta de um certo “parentesco”, podem ser entendidos como linguagem.⁷⁸ Por ser uma instituição social, a linguagem faz “parte de de uma atividade ou de forma de vida”⁷⁹, de maneira que retirá-la desse contexto impossibilita a formação de sentido. Nesse caso, a “linguagem sai de folga” e por conta disso é que surgem os problemas filosóficos.⁸⁰

Pode-se afirmar, portanto, que o nível da semântica está adstrito ao da pragmática, pois o sentido de palavras e frases só se dá quando elas são utilizadas de acordo com as regras de um jogo de linguagem específico. Isso implica que o uso correto de um termo pressupõe a capacidade de seguir regras ou “agir dentro de uma dada situação linguística”.

Por “uso correto” não se entende que há uma única forma de se utilizar o termo, mas que há “casos normais” e “anormais” de aplicação, à medida que o uso de um termo ocorre sem dificuldades ou em situações em que não se sabe exatamente como proceder. Dessa forma, jogos de linguagem mais simples ou

⁷⁸ Cf. WITTGENSTEIN, L., *Investigações Filosóficas*, §65 e SIMON, Henrique Smidt, op. cit., p. 57.

⁷⁹ WITTGENSTEIN, L., *Investigações Filosóficas*, §23.

⁸⁰ WITTGENSTEIN, L., *Investigações Filosóficas*, §38.

primitivos funcionam como parâmetro para a compreensão de jogos mais complexos, o que se torna possível por conta de suas “semelhanças de família”.⁸¹

Seguir regras e dominar jogos de linguagem são processos bastante assemelhados. Em ambos, presume-se a existência de um padrão ou modelo de atuação, que deve ser público e reiterado, e é o afastamento desse modelo generalizado que cria a possibilidade de existência de um erro particular.

Tais regras, contudo, são aplicadas seguindo um critério de familiaridade e não de precisão, de maneira que funcionam apenas como uma “placa de sinalização” para o caminho a ser seguido. A linguagem é entendida, destarte, como um “acordo de modos de vida dos homens” e as regras de uso das palavras são, na verdade, padrões de significabilidade.⁸² Sob essa ótica, a função da filosofia passa a ser entendida como uma “terapia” de esclarecimento.⁸³

Esse modo de concepção da linguagem, aliado às teorias de John L. Austin (especialmente seu conceito de *atos de fala*), influenciou Herbert Hart, em sua obra *O Conceito de Direito*, na qual ele concentra a ênfase de sua análise nos contextos e nos usos imprecisos da linguagem. Assim, Hart formula um conceito “aberto” de direito, que “não se pretende preciso e determinado”, mas que estabelece padrões de certos elementos que, através de critérios de “semelhanças de família”, permitem denominar algo como “direito”, diferenciando-o de outras instituições sociais normativas.⁸⁴

De acordo com essa visão, o direito, assim como a linguagem para o Wittgenstein das Investigações Filosóficas, é entendido como um fenômeno social e deixa-se de buscar por uma “essência” que seja capaz de responder a pergunta “o que é direito?”. Avança Hart que sistemas de regras possuem um aspecto interno e outro externo, utilizando como ilustração de seu raciocínio uma partida de xadrez.⁸⁵

Os padrões de movimento das peças dos jogadores não são apenas executados com uma certa regularidade, mas também são vistos por eles como “um padrão para tantos quantos pratiquem o jogo”, como se tivessem “uma opinião

⁸¹ SIMON, Henrique Smidt, op. cit., pp.58-60.

⁸² SIMON, Henrique Smidt, op. cit., pp. 63-65, 100; WITTGENSTEIN, L., Investigações Filosóficas, §§65-68.

⁸³ SIMON, Henrique Smidt, op. cit., p. 118.

⁸⁴ SIMON, Henrique Smidt, op. cit., p. 101.

⁸⁵ HART, Herbert L. A.. O Conceito de Direito, pp. 65.

formada” sobre a correção daquele movimento.⁸⁶ Tal opinião se manifesta através de uma “terminologia normativa” que formula, por exemplo, críticas, reconhecimentos e exigências acerca de tais regras, especialmente quando são desrespeitadas, do gênero: “você (não) deve fazer isso...”.⁸⁷

Entretanto, um observador externo apenas pode verificar a existência de certos padrões de ação e de uma consequência em caso de sua quebra, o que pode levá-lo a entender um comportamento espontâneo como uma regra e vice-versa. Assim, para compreender o modo como os indivíduos experienciam as regras, é necessário, além de observar, ser capaz de participar do jogo, razão pela qual se diz que essa perspectiva privilegia o sujeito que faz parte do sistema, aquele que “vivencia o conjunto de regras”, ou seja, enfatiza seu aspecto interno.⁸⁸

O direito é caracterizado, dessa forma, por meio do compartilhamento de um paradigma em comum entre os indivíduos, que os permite identificar padrões de comportamento, através de semelhanças de família, que fazem com que certas regras sejam aceitas como direito. Ainda assim, para que se diferencie o direito de outras práticas sociais normativas, é necessário definir certos critérios.

Para tanto, Hart defende que o direito pode ser entendido como um sistema em que normas primárias e secundárias se complementam.⁸⁹ As primeiras são aquelas que determinam uma ação ou abstenção por parte das pessoas, enquanto as segundas estabelecem formas de criar, extinguir, modificar regras primárias, bem como fiscalizar a aplicação destas.⁹⁰

A introdução de tais regras em um sistema permite tratar questões como a incerteza acerca da validade de uma norma, a flexibilidade do próprio sistema e aplicação de uma sanção em caso de transgressão. Por conta disso, regras secundárias têm papel fundamental na “identificação do direito como um jogo de linguagem diferente da obediência aos costumes ou às regras de cortesia”.⁹¹

Dadas tais funções, as regras secundárias podem ser subdivididas em três categorias: alteração, julgamento e reconhecimento. As regras de reconhecimento

⁸⁶ HART, Herbert L. A.. O Conceito de Direito, pp. 65-66.

⁸⁷ Cf. HART, Herbert L. A., op. cit., p. 66 e SIMON, Henrique Smidt, op. cit., p. 103.

⁸⁸ SIMON, Henrique Smidt, op. cit., pp. 99,103.

⁸⁹ HART, Herbert L. A.. O Conceito de Direito, pp. 89 e ss..

⁹⁰ SIMON, Henrique Smidt, op. cit., pp. 106-107.

⁹¹ SIMON, Henrique Smidt, op. cit., p. 107.

são de especial importância, pois permitem definir a validade de uma regra e seu âmbito de aplicação. Regras de reconhecimento, geralmente, não são enunciadas, mas apenas se manifestam, ou seja, sua existência é uma questão de fato.

Por conta disso, a regra de reconhecimento possui dois aspectos, a saber, um de direito e outro de fato, pois manifesta-se tanto na visão interna do sistema, quanto na visão externa. Na perspectiva interna, é ela que permite ao participante do jogo reconhecer uma regra como vinculativa (ainda que não concorde com ela), ao mesmo tempo em que sua existência somente se verifica a partir de uma análise externa dos padrões comportamentais complexos do sistema.⁹²

Por conta disso, às regras de reconhecimento não se aplicam critérios de validade, pois elas são o próprio fundamento de validade do sistema. Como dito, o direito, assim como um jogo de linguagem, se dá através de uma forma de vida, de forma que a regra de reconhecimento simplesmente manifesta acordos intersubjetivamente compartilhados como padrões de conduta que devem ser seguidos. Como dito, trata-se de uma questão de fato, pois os indivíduos “simplesmente agem dessa forma”.⁹³

Para que um sistema jurídico exista, é necessário que as regras primárias sejam obedecidas com certa regularidade, ao passo que as regras secundárias devem ser aceitas, ao menos por aqueles que aplicam o direito, como “padrões públicos e comuns de comportamento oficial”.⁹⁴ Não se trata, todavia, de uma definição rígida de direito, pois o objetivo da obra de Hart é justamente uma concepção (aberta) que assuma o direito como uma prática social que, como tal, é muitas vezes dotada de vagueza e imprecisão.

Uma “definição” dessa natureza permite a conclusão de que haverá casos de aplicação do direito que não possuirão uma solução única. Isso porque, sob essa perspectiva, o direito pode ser entendido como um jogo de linguagem, cujas regras não são fixas e, portanto, podem gerar dúvidas sobre como proceder.

Ademais, ao passo que o direito utiliza a linguagem ordinária para formular seus conceitos, ele também é influenciado pelos modos de vida que são a ela

⁹² SIMON, Henrique Smidt, op. cit., pp. 109-110.

⁹³ SIMON, Henrique Smidt, op. cit., p. 110.

⁹⁴ HART, Herbert L. A., op. cit., p. 128.

inerentes. Assim, se o contexto herdado é alterado, também mudarão as vivências do destinatário da norma e, portanto, suas percepções sobre o fenômeno jurídico.⁹⁵

2.3 O decisionismo dentro do paradigma da filosofia analítica

Uma vez que regras de direito possuem seres humanos como destinatários, torna-se imperioso que seus conceitos e ordens sejam-lhes inteligíveis. Dessa forma, o direito torna-se uma espécie de “refém” da linguagem natural, o que leva ao problema da imprecisão: como proceder quando se tem dúvida quanto à aplicação de uma norma a um caso prático? Kelsen e Hart abordam o problema formas distintas, uma vez que suas concepções acerca dos fenômenos linguístico e jurídico são bastante diferenciadas.

Ainda assim, dados os pressupostos do paradigma da filosofia analítica da linguagem, um problema sempre persistirá: haverá situações em que os elementos de uma norma possuirão mais de um significado ou gerarão dúvida quanto à sua subsunção, possibilitando decisões divergentes, porém juridicamente válidas. Ocorre que tanto Kelsen, quanto Hart, separam o direito da moral e entendem que a norma jurídica independe de seu conteúdo, negando um critério valorativo único, capaz de definir de antemão uma solução correta dentro das várias possíveis.

Por consequência, não se pode falar em um direito justo ou em um critério de interpretação universalmente válido, de forma que haverá casos em que a aplicação do direito pelo magistrado dependerá de juízos de conveniência e oportunidade. A questão do decisionismo pode ser formulada como o “reconhecimento de uma margem de discricionariedade na aplicação do direito pelo magistrado, de modo que a teoria não possa dizer *a priori* qual interpretação é melhor que a outra”⁹⁶. Para melhor se compreender essa problemática, cumpre distinguir alguns dos modos de incertezas significativas da linguagem.

Dá-se significado a um termo através de características designativas e denotativas. A designação se refere aos atributos que podem ser derivados de um termo, ou seja, uma espécie de descrição das características que um certo exemplar

⁹⁵ WITTGENSTEIN, L., *Investigações Filosóficas*, §23.

⁹⁶ SIMON, Henrique Smidt, *op. cit.*, p. 17.

deve possuir para que possa ser incluído no rol de coisas referentes àquele termo. Por denotar, entende-se explicitar ou enumerar os exemplares que podem integrar a extensão de um termo.⁹⁷

Por conta disso, tem-se que a vagueza é um atributo essencialmente denotativo, enquanto a ambiguidade é predominantemente designativa. Com relação à primeira, pode-se dizer que um termo possui três zonas: *luminosidade positiva* (casos que estão com certeza dentro de sua denotação), *luminosidade negativa* (casos que certamente não estão em sua denotação) e *incerteza* (situações em que há dúvidas legítimas quanto à sua extensão).

É na zona de incerteza que ocorre a vagueza, ou seja, “quando, a partir das regras de uso, não se pode definir com exatidão os critérios de aplicabilidade denotativa, devendo-se recorrer a decisões auxiliares”. A inclusão ou não das situações incertas na extensão de um termo é, portanto, uma decisão convencional dos usuários da linguagem.

Por outro lado, quando regras de uso geram incerteza designativa, está-se diante da ambiguidade, entendida como quando “um mesmo rótulo ou significante veicula propriedades designativas aplicáveis a âmbitos denotativos claramente diferenciáveis”.⁹⁸ No campo do direito, frequentemente, recusa-se a possibilidade de existência tanto de vaguezas, quanto de ambiguidades, nas normas.⁹⁹

As teorias de Kelsen sobre a interpretação se dão sob os pressupostos epistemológicos do Neopositivismo, que estabelecem funções específicas às definições linguísticas nas tarefas científicas. Assim, ao mesmo tempo em que os neopositivistas objetivam eliminar definições axiológicas (metafísicas), eles também procuram maior precisão nas definições de conceitos. Todavia, as definições da linguagem natural são vagas, de maneira que se torna necessário elucidá-las (torná-las mais precisas), para que se possa “decidir, com relativo grau de consistência, se um objeto pode ou não ser subsumido sob um dado conceito”.¹⁰⁰

Uma das funções das normas é dotar de sentido objetivo uma conduta social. Para tanto, um órgão competente deve interpretá-la de modo a definir a apreensão

⁹⁷ WARAT, op. cit., p. 55.

⁹⁸ WARAT, op. cit., p. 78.

⁹⁹ WARAT, op. cit., pp. 76-79.

¹⁰⁰ WARAT, op. cit., p. 59.

dos seus diversos significados possíveis. Na perspectiva da Teoria pura, a interpretação de normas consiste em duas atividades: um juízo de adequação da norma hierarquicamente inferior com relação à superior e na delimitação dos possíveis significados de seus termos.¹⁰¹

O processo interpretativo ocorre em diversos momentos, tais como na determinação do conteúdo da norma pelos órgãos encarregados de aplicá-la, na definição de seu sentido pelos cientistas do direito ou, ainda, quando os próprios indivíduos pautam suas condutas com base em seus entendimentos sobre uma determinada regra de conduta.

Todavia, somente a atividade interpretativa realizada pelos tribunais é dotada de força prescritiva, o que configura a interpretação autêntica, que se contrapõe à não-autêntica. A interpretação é necessária porque a norma jurídica pode ser alvo de indeterminações, sejam elas intencionais ou não. Uma indeterminação intencional decorre de uma margem de discricionariedade que uma norma mais geral confere a uma menos geral para garantir maior eficiência em sua aplicação, que depende, portanto, de juízos de oportunidade diante das circunstâncias do caso concreto.¹⁰²

Já a determinação não-intencional decorre dos próprios limites da linguagem natural, que pode ser ambígua ou excessivamente vaga, ou da preceituação de condutas contraditórias por normas de mesma hierarquia. Nos casos de vagueza e ambiguidade, cabe à ciência jurídica apenas delimitar os possíveis significados da norma, criando um tipo de “moldura” que estabelece as possíveis interpretações, dentre as quais o órgão aplicador do direito terá liberdade para decidir qual utilizar. Independentemente da solução escolhida, desde que ela esteja dentro da moldura delimitada pela ciência jurídica, a decisão ocorrerá conforme a lei, pois pode-se entender que o próprio legislador atribuiu vários sentidos à norma e que tal escolha é de natureza jurídico-política.¹⁰³

Ainda assim, mesmo que o juiz escolha um sentido que exceda os limites da moldura interpretativa (fora, portanto, dos limites da lei), ao transitar em julgado, sua decisão não estará ferindo o ordenamento jurídico, pois este lhe confere a

¹⁰¹ SIMON, Henrique Smidt, op. cit., p. 125.

¹⁰² SIMON, Henrique Smidt, op. cit., pp.122-123.

¹⁰³ SIMON, Henrique Smidt, op. cit., pp. 119-122.

oportunidade de dar à norma seu próprio entendimento, o que caracteriza o decisionismo.¹⁰⁴

Para Hart, o problema do decisionismo adquire novos contornos, uma vez que leva em conta a complexidade da linguagem natural. Nesse caso, o fato de sua perspectiva assimilar o âmbito pragmático da linguagem leva à consequência de que haverá casos em que a aplicação da norma se afasta da regra padrão, gerando incerteza, de modo que será necessário, dentro de certos limites, definir a extensão de conceitos normativos. Nas palavras de Hart:¹⁰⁵

Boa parte da teoria do direito deste século tem-se caracterizado pela tomada de consciência progressiva (e, algumas vezes, pelo exagero) do importante facto de que a distinção entre as incertezas da comunicação por exemplos dotados de autoridade (precedente) e as certezas de comunicação através da linguagem geral dotada de autoridade (legislação) é de longe menos firme do que sugere este contraste ingénuo. Mesmo quando são usadas regras gerais formuladas verbalmente, podem, em casos particulares concretos, surgir incertezas quanto à forma de comportamento exigido por elas.

Mais uma vez, a linguagem representa um limite à aplicação das regras (assim como em todos os campos da experiência humana). Há “casos simples”, que ocorrem em contextos semelhantes, para os quais aplicar a regra será uma questão de mero silogismo, mas também há situações em que isso não ocorrerá de forma automática. Todo caso que fugir daquele tipificado na norma como regra geral entra na “zona cinzenta” de aplicação do dispositivo normativo e deverá ser interpretado. Nesse caso, cabe ao intérprete verificar se a situação em questão pode ser acrescentada aos casos de aplicação da norma, por ser “juridicamente relevante” ou “suficientemente próxima”.¹⁰⁶

Tal processo, evidentemente, acarreta uma atividade decisória por parte da pessoa que aplica a norma, pois “o poder discricionário que assim lhe é deixado pela linguagem pode ser muito amplo; de tal forma que, se ela aplicar a regra, a conclusão constitui na verdade uma escolha, ainda que possa não ser arbitrária ou

¹⁰⁴ SIMON, Henrique Smidt, op. cit., p. 124.

¹⁰⁵ HART, Herbert L. A., op. cit., p. 139.

¹⁰⁶ HART, Herbert L. A., op. cit., p. 140.

irracional.”¹⁰⁷ Essas indeterminações compõem a zona de *textura aberta* do direito e, frise-se, são consequências da utilização da linguagem humana.

Todavia, não se deve ter por ideal um ordenamento jurídico que defina previamente todos os casos de aplicação de suas normas, pois, além de impossível, isso tampouco é desejável. Acontece que sistemas normativos terão sempre de contrabalancear duas necessidades: permitir que os indivíduos apliquem com segurança regras a si mesmos e deixar em aberto situações que só podem ser resolvidas por órgãos oficiais diante de um caso concreto.¹⁰⁸

Ainda assim, há um limite para a discricionariedade, pois o juiz deverá manter-se tanto quanto possível próximo dos usos já estabelecidos dos termos em questão. Os casos padrões funcionam como instâncias públicas de controle que, por meio de semelhanças de família, tornam possível sua aplicação a novos casos não previstos pela norma, característica que aproxima a *textura aberta* de Hart com o problema de seguir regras de Wittgenstein.¹⁰⁹

Isso posto, pode-se constatar que, enquanto a problemática das definições enfatiza, para o positivismo lógico, a estipulação de critérios rígidos na interpretação de enunciados para empregá-los na atividade científica, na filosofia da linguagem ordinária, há uma preocupação com a clarificação do sentido léxico dos termos. No campo do direito, isso se reflete principalmente no que diz respeito à interpretação de leis.

Nesse sentido, explica Warat¹¹⁰ que, por meio da análise dos contextos em que os termos são empregados, criam-se definições de domínio, de modo que são detectadas semelhanças de família entre os diversos significados. Para tanto, é necessário “estabelecer-se de forma mais ou menos arbitrária um significado central definido através de uma série de características designativas”¹¹¹, ao passo que a alteração de algumas dessas características criam variantes do caso paradigmático.

Nesse sentido, entende-se que a interpretação de leis leva a um processo redefinitório, seja ele direto ou indireto. Redefinir consiste em alterar características

¹⁰⁷ HART, Herbert L. A., op. cit., p. 140.

¹⁰⁸ HART, Herbert L. A., op. cit., p. 143.

¹⁰⁹ SIMON, Henrique Smidt, op. cit., pp. 126-128.

¹¹⁰ WARAT, op. cit., pp. 79-80.

¹¹¹ WARAT, op. cit., p. 79.

designativas para permitir mudanças no campo denotativo, processo que, no poder judiciário, é sempre axiológica ou ideologicamente determinado.¹¹² Isso pode ocorrer tanto por conta da tentativa de aclarar termos ambíguos ou vagos presentes nas leis (redefinições diretas), quanto através de recursos retóricos que objetivam mudanças designativas desses termos, escamoteando aspectos extratextuais das atividades interpretativas (redefinições indiretas).

Por fim, a questão da “interpretação criadora”, que persiste tanto no neopositivismo, quanto na filosofia da linguagem ordinária, pode ser analisada como um certo traço presente no paradigma da filosofia analítica, no que tange à questão dos significados. Nesse sentido, uma das características peculiares ao modo analítico de se pensar a linguagem, segundo Cabrera, é uma espécie de aversão aos significados enquanto entidades substantivas, o que ele denominou de “Tese S”:

[Sob a perspectiva das filosofias analíticas] os "significados" são entidades problemáticas (decididamente obscuras em algumas de suas versões), de difícil acesso e identificação, cujo valor explicativo é duvidoso e cujo ônus metafísico é elevado. Semânticas que deixam totalmente de lado os “significados”, ou aquelas que tentam elucidá-los em termos referencialistas ou em termos de usos da linguagem, parecem ter melhores chances de formulação científica ou tratamento filosófico.¹¹³

Dessa forma, tanto os neopositivistas, quanto os filósofos da linguagem ordinária, acabam por chegar em um tipo de impasse com relação ao significados. Para ambos, ao se evitar pressupostos ditos “metafísicos” ou pré-concepções subjetivas de mundo, torna-se contraditório afirmar que os termos possuem significados previamente definíveis.

Como consequência, termos possuirão mais de um significado possível, podendo, no máximo, serem estabelecidas barreiras interpretativas que definam os limites das várias significações possíveis. Nesse sentido, é ainda mais plausível a hipótese de que a linguagem consiste numa atividade que não simplesmente veicula pensamentos, mas também constitui ideias à medida que elas são enunciadas.

¹¹² WARAT, op. cit., p. 79.

¹¹³ CABRERA, op. cit., p. 45.

Capítulo 3 - Artificialmente parcial: contexto e vieses em algoritmos

Visto sob um ponto de vista da filosofia da linguagem ordinária, o processamento de linguagem natural pode ser alvo de novas indagações. Como pode, afinal, um algoritmo que opera através de uma linguagem computacional, estritamente lógica, ser capaz de “compreender”, ainda que minimamente, a linguagem humana? Do ponto de vista da semiótica, técnicas de NLP estariam tentando levar a cabo a assimilação de uma linguagem que opera em três níveis: sintático, semântico e pragmático, ao passo que a linguagem por meio da qual seus algoritmos foram escritos está limitada aos dois primeiros.

Certamente a resposta para tal questionamento é complexa e levará em conta não apenas os questionamentos da filosofia da linguagem, mas também da linguística, ciências da computação, psicologia cognitiva, dentre outros campos do conhecimento. Ainda assim, para os fins deste trabalho, a perspectiva filosófica é suficiente para apontar a relevância do estudo da linguagem humana para que melhor se possa compreender os desafios que envolvem a elaboração de algoritmos de NLP.

Dadas as reflexões até aqui expostas, é possível afirmar que uma compreensão plena da linguagem excederia as competências de uma inteligência artificial de propósito limitado (fraca), dado que o comunicar humano envolve operações bastante complexas, tais como abstrações, generalizações, bem como certas pré-concepções acerca dos falantes e do ambiente no qual eles estão inseridos. Tais habilidades, inerentes à inteligência humana, só seriam possíveis de serem desempenhadas por uma inteligência artificial de propósito geral.

De todo modo, entre a assimilação plena da linguagem natural e sua total incompreensão, há diversas situações intermediárias, nas quais se situa o atual estágio de desenvolvimento do processamento de linguagem natural. No que tange ao campo pragmático da linguagem, algoritmos são capazes de utilizar técnicas matemáticas para “calcular” o significado de termos a partir de sua ocorrência em determinados “contextos” textuais. A ideia subjacente a esse método é que o agregado de contextos nos quais um termo ocorre é capaz de definir características

que revelam similaridades de significado entre palavras e conjuntos de palavras entre si.¹¹⁴

No capítulo primeiro deste trabalho, abordou-se a questão da vetorização de textos e como tal processo é utilizado para, dentre outras funções, classificar documentos. Todavia, resta explorar, ainda que sucintamente, como o aprendizado de máquina trata da questão de dar significado a diferentes termos, tarefa que permite, por exemplo, identificar relações semânticas e sintáticas, expandir conceitos e identificar estruturas argumentativas em documentos.¹¹⁵

Defender-se-á que técnicas de NLP que dão maior ênfase ao contexto textual em que os termos se encontram alcançam maior êxito na “compreensão” de seus significados. De todo modo, não parece possível equiparar tais processos àquilo que é comumente tido como entendimento humano da linguagem. Ainda assim, isso não é necessário para que tais algoritmos sejam capazes de um desempenho bastante satisfatório em tarefas que envolvem o processamento de linguagem natural, trazendo consigo, destarte, diversas possibilidades de aplicação dentro e fora do campo jurídico.

3.1 Ensinando a máquinas o contexto

É forçoso dizer que foi a leitura de textos sobre filosofia da linguagem por parte de cientistas da computação que permitiu avanços na área do processamento de linguagem natural, mas é fato que uma mudança de abordagem, através da qual se tem uma maior preocupação com o contexto dos termos, possibilitou uma efetiva melhora no desempenho desse tipo de algoritmo. Por conta de tais mudanças, há quem afirme a ocorrência de uma “virada pragmática” no processamento de linguagem natural, de modo que algoritmos como o Word2vec seriam uma “prova empírica” de que as ideias de Wittgenstein sobre jogos de linguagem estariam

¹¹⁴ASHLEY, Kevin D., op. cit., p. 242.

¹¹⁵ Cf. item 1.2 deste trabalho.

corretas, refutando a ideia de que palavras são entidades isoladas, signos que simplesmente representam um certo objeto existente no mundo real.¹¹⁶

O conjunto de técnicas de NLP consistentes em adaptar palavras e frases em vetores matemáticos dotados de representações numéricas constitui o *word embedding*.¹¹⁷ Acontece que redes neurais podem “aprender” somente a partir de dados numéricos, o que implica a necessidade de converter dados textuais em entidades matematicamente analisáveis, o que pode ser feito de diferentes formas. A função do *word embedding* é facilitar o aprendizado de redes neurais ao representar referidos dados de forma mais eficiente, o que faz através de vetores.¹¹⁸ Todavia, ao criar tais representações, esses modelos permitem igualmente a identificação de novos tipos de relações entre as palavras em um determinado vocabulário.

Uma abordagem “clássica” para lidar com dados textuais, o Bag-of-Words (BoW), consiste em transformar cada palavra de um vocabulário de N termos em um vetor de tamanho N (chamado de vetor *one-hot*), composto apenas de “zeros”, exceto por um “um”, que permite diferenciá-lo dos demais termos. Assim, um vocabulário de três termos (“rei”, “rainha” e “carro”) poderia ser representado como: “rei [0,0,1]”, “rainha [0,1,0]” e “carro [1,0,0]”. Através do BoW, documentos inteiros também são representados por um vetor único, no qual cada um de seus elementos representa o número de ocorrências (ou não ocorrência) do enésimo termo naquele documento, de forma que tais dados podem ser utilizados por classificadores que levam em conta a ocorrência de palavras como *features*.

Acontece que, em tal abordagem, além da ineficiência em termos computacionais, uma vez que é necessário realizar operações matemáticas com entidades compostas essencialmente por zeros, cada palavra é tratada como apenas uma dimensão em um espaço vetorial, que não interfere nas demais, ou

¹¹⁶ BELLONI, Massimo. Neural Networks and Philosophy of Language. Disponível em: <<https://towardsdatascience.com/neural-networks-and-philosophy-of-language-31c34c0796da>>. Acesso em: 16 nov. 2019.

¹¹⁷ A tradução literal de *word embedding* para o português seria algo como “embutimento” ou “acondicionamento de palavras”. Todavia, trata-se de um termo que é majoritariamente utilizado em inglês, motivo pelo qual optou-se por utilizá-lo na forma original neste trabalho.

¹¹⁸ KARANI, Dhruvil. Introduction to Word Embedding and Word2Vec. 2018. Disponível em: <<https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>>. Acesso em: 16 nov. 2019.

seja, todas são igualmente distantes entre si. Dessa forma, palavras como “rei” e “rainha” são tão diferentes entre si quanto são da palavra “carro”. É bastante evidente, contudo, para um humano, que as duas primeiras palavras guardam bastante similaridade, distinguindo-se apenas quanto ao gênero, e que ambas guardam certa distância de “carro”, ao passo que estão igualmente próximas de “palácio”.

Ainda assim, não é fácil explicar tais relações para uma máquina. Na verdade, o que um humano quer dizer com palavras “parecidas” ou “diferentes” significa que certos termos possuem maior chance de ocorrer em um determinado contexto do que em outros. Dessa forma, ao falar-se em “casa”, “quintal” e “jardim”, as palavras “cachorro” e “gato” terão maior número de ocorrências que, por exemplo, “elefante”, ainda que todas se refiram a mamíferos.

Nesse sentido, o Word2vec, uma outra abordagem de *word embedding*, desenvolvida em 2013, por um time de engenheiros da Google, liderados por Tomas Mikolov, buscou trazer maior eficiência a esse processo. Ao contrário de modelos que tratam palavras como “unidades atômicas” e que não possuem, portanto, uma noção de “similaridade”, o Word2vec estabelece um modelo mais complexo, que cria representações para cada um dos termos de seu vocabulário analisando as demais palavras que compõem o seu contexto.

Essa mudança de abordagem permite que termos sejam representados em um número de dimensões bastante inferior ao número N de palavras que compõem o *corpus*, o que aumenta sua eficiência computacional e permite que o modelo seja treinado em conjuntos de dados milhares de vezes maiores que seus antecessores.

O Word2vec faz isso ao treinar de forma não supervisionada uma rede neural que tem por objetivo adivinhar uma determinada palavra com base em um conjunto de palavras contexto (*Continuous Bag of Words - CBOW*) ou realizando o processo inverso (*Skip-Gram*).¹¹⁹ Assim, a rede neural cria exemplos positivos e negativos a partir de frases do *corpus* (e.g. “o cão está deitado no jardim”), substituindo a palavra-alvo (*target word*) por palavras aleatórias daquela base de dados (e.g. “o golfinho está deitado no jardim” ou “o casa está deitado no jardim”), o que a possibilita criar dados de treino e de teste e aprimorar suas previsões.

¹¹⁹ MIKOLOV, Tomas et al. Efficient Estimation of Word Representations in Vector Space, pp. 3-4.

Por conta dessa metodologia, é possível que a rede neural posicione em regiões próximas de um espaço vetorial termos que normalmente são acompanhados por palavras-contexto similares e distantes de termos que quase nunca ocorrem juntos daquelas. Na prática, isso significa, que o modelo “compreendeu” a existência de certas relações sintáticas e semânticas entre os termos do vocabulário. Não apenas isso, torna-se possível realizar, inclusive, operações algébricas com palavras.¹²⁰

Por exemplo, podem-se distinguir diferentes tipos de relações entre palavras ao somar ou subtrair os vetores que as representam. Sabe-se que “a palavra *grande* é similar a *maior* no mesmo sentido que *pequeno* é para *menor*”, de maneira que se pode explicitar tais relações através de perguntas, que podem ser formuladas em termos matemáticos, tais como “qual palavra é semelhante a *pequeno* no mesmo sentido que *o maior* é similar a *grande*?¹²¹. Seguindo a mesma lógica, chega-se a “Roma” através da operação “Paris - França + Itália”¹²², o que significa que Paris está para França da mesma forma que Roma está para Itália.

¹²⁰ MIKOLOV, op. cit., p. 2.

¹²¹ Tradução livre: “the word big is similar to bigger in the same sense that small is similar to smaller” e “What is the word that is similar to small in the same sense as biggest is similar to big?” em MIKOLOV, op. cit., p. 5.

¹²² A ideia é estimar o valor numérico da relação “ser a capital de” ao se extrair do vetor “Paris” o vetor “França” e adicionar tal valor ao vetor “Itália”, o que leva a algo próximo de “Roma”. O mesmo raciocínio se aplica a “Rei - Homem + Mulher = Rainha” ou “Grande - Maior + Pequeno = Menor”.

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

Figura 6 - Exemplos de cinco tipos de relações semânticas e nove tipos de relações sintáticas encontradas pelo Word2Vec.¹²³

Dentre as contribuições de modelos que são capazes de assimilar tais relações linguísticas, figuram o aprimoramento de aplicações atuais de NLP, tais como traduções por máquina, *information retrieval* e *question answering systems*, bem como viabilizar que novas técnicas sejam criadas.¹²⁴

Após o Word2vec, foram propostas melhorias no *word embedding*, tais como o GloVe (*Global Vectors for Word Representation*)¹²⁵, da universidade de Stanford e o fastText, do Facebook. O primeiro deles aponta que métodos que criam vetores de palavras a partir de “janelas de contexto locais”, sem analisar um documento como um todo (como é o caso do Word2vec), “conseguem captar analogias entre palavras, mas não utilizam de forma satisfatória as relações estatísticas do *corpus*”.¹²⁶ Para superar essa limitação, os criadores do GloVe propõem uma matriz global que relacione as co-ocorrências de cada um dos termos do *corpus*, o que, além de uma

¹²³ MIKOLOV, op. cit., p. 6.

¹²⁴ MIKOLOV, op. cit., p. 5; Cf. Seção 1.2 deste trabalho.

¹²⁵ Cf. PENNINGTON, Jeffrey; SOCHER, Richard; MANNING, Christopher. Glove: Global Vectors for Word Representation. Proceedings Of The 2014 Conference On Empirical Methods In Natural Language Processing (emnlp).

¹²⁶ PENNINGTON, Jeffrey, op. cit., p. 2.

melhoria de performance, permite entender melhor as origens das regularidades semânticas e sintáticas da representação de palavras.¹²⁷

Outro tipo de limitação do Word2vec, segundo os criadores do fastText, é que aquela abordagem não leva em conta as estruturas internas das palavras. Assim, ao invés de definir cada termo como um vetor *one-hot*, o fastText cria n-gramas, que possibilitam representar uma palavra de diferentes formas, de modo que é possível relacionar n-gramas de diferentes palavras entre si, bem como criar vetores de palavras que não constavam originalmente no *corpus*.¹²⁸

Por fim, ainda que o *word embedding* represente um grande avanço no processamento de linguagem natural, deve-se fazer a ressalva de que tais tecnologias ainda estão aquém do nível de compreensão humano da linguagem. Captar o contexto textual das palavras não necessariamente significa entender a linguagem como uma forma de vida, tampouco ser capaz de “jogar seus jogos”.

A importância do contexto, conforme explorado no capítulo anterior, não diz respeito somente a co-ocorrências de palavras, mas também às próprias vivências coletivas dos usuários da linguagem. A representação vetorial de palavras estaria mais próxima da atitude de um observador externo, que apenas descreve regularidades que ele verifica ao analisar a transcrição de certos jogos.

Dessa forma, um dos argumentos que serão sustentados neste trabalho é que os algoritmos, por não serem capazes de compreender sozinhos a linguagem natural, terão de “importar” dos seres humanos certas pré-compreensões, que estarão presentes nos dados que são através dos quais eles são treinados. Esse processo, todavia, pode ser falho e resultar no ocultamento de certas opiniões, muitas vezes preconceituosas.

3.2 Vieses: como opiniões se tornam “ciência”

3.2.1 Heurística e vieses

¹²⁷ PENNINGTON, Jeffrey, op. cit., p. 1.

¹²⁸ BELLONI, Massimo. Neural Networks and Philosophy of Language. Disponível em: <<https://towardsdatascience.com/neural-networks-and-philosophy-of-language-31c34c0796da>>. Acesso em: 16 nov. 2019.

Preconceito e resistência parecem ser mais a regra do que exceção no desenvolvimento científico avançado. Além disso, em condições normais eles caracterizam a melhor investigação, a mais criativa e também a mais rotineira. Não está também em questão qual a sua origem. Não se tratam de características anômalas de indivíduos, mas de características da comunidade com raízes profundas no processo como os cientistas são treinados para trabalhar na sua profissão.¹²⁹

Antes de se explorar a questão da pressuposição de neutralidade das máquinas, cumpre abordar, ainda que brevemente, como as decisões humanas estão sujeitas a heurísticas e vieses. Nesse sentido, a psicologia cognitiva trouxe importantes contribuições, trazendo à tona algumas das possibilidades de engano na cognição humana e, ainda que seja incipiente a literatura específica sobre vieses decisórios no direito¹³⁰, não há motivos para crer que o processo de decisão judicial não esteja submetido às mesmas limitações de outros tipos de decisões.¹³¹

Partindo-se de modelos teóricos que distinguem dois modos coexistentes de pensar, também conhecidos por *teorias do processo dual*, dividem-se as atividades cognitivas humanas em dois sistemas: um mais rápido e intuitivo (apelidado de Sistema 1) e outro deliberado e devagar (Sistema 2).¹³² Ambos os sistemas (conjuntos de processos mentais) atuam concorrentemente, de forma que não podem ser entendidos como unidades autônomas, mas com características bastante distintas no que diz respeito à velocidade, controlabilidade e conteúdo de suas operações.¹³³

O Sistema 1 serve para tomar decisões rápidas, com pouco esforço cognitivo e quando se dispõe de pouca informação, estando comumente associado a processos intuitivos ou “natos”. Especialmente em sociedades primitivas, nas quais

¹²⁹ KUHN, Thomas. A função do dogma na investigação científica, p. 24.

¹³⁰ No direito brasileiro, exploram esse tema Alexandre Morais da Rosa, em Guia do Processo Penal Conforme a Teoria dos Jogos; Paola Wojciechowski, em Vieses da justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva e; Dierle Nunes et al, em Desconfiando da imparcialidade dos sujeitos processuais.

¹³¹ WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. Vieses da justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva, p. 48.

¹³² Cf. KAHNEMAN, Daniel. Rápido e devagar: duas formas de pensar; WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., p. 22.

¹³³ WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., p. 23.

uma resposta rápida a um estímulo ambiental poderia significar a diferença entre a vida e a morte, ele foi (e continua sendo) bastante útil.

Todavia, em sociedades “desenvolvidas”, não levar em conta sua influência nos processos decisórios pode levar a erros sistemáticos de pensamento.¹³⁴ O Sistema 1 atua por meio de atalhos cognitivos, também denominados de heurísticas, que nada mais são que mecanismos de reconhecimento de informações, que ajudam a encontrar respostas simples, ainda que imperfeitas, para perguntas que demandariam maior reflexão e esforço cognitivo - características do Sistema 2.¹³⁵

Assim, partindo do acúmulo de informações pretéritas, o Sistema 1 responde automaticamente em caso de ocorrência de uma situação semelhante.¹³⁶ Na verdade, esse sistema está em busca de coerência, de modo que ele acaba, não raro, por confundir “familiaridade e conforto cognitivo com ilusões de veracidade”, tornando-se insensível “à quantidade e à qualidade da informação que estiver na gênese das impressões e intuições”. Dessa forma, a busca por consistência das histórias narradas pelo Sistema 1 é muitas vezes gerada sem levar em conta a deficiência da informação disponível.¹³⁷

O Sistema 2, por outro lado, exige esforço e atenção e está relacionado a processos lógicos, sequenciais, e conscientes, mas também lentos e ineficientes. Além disso, ele é o único que permite pensar abstratamente e “programar a memória para obedecer instruções que encarnam tarefas não habituais”¹³⁸, como, por exemplo, encontrar objetos de características deliberadamente escolhidas em meio a vários outros. Além disso, a repetição de tarefas que inicialmente demandavam o Sistema 2, tais como dirigir ou praticar algum esporte, acaba por transferi-las ao Sistema 1, permitindo sua execução de forma automática.

Os dois sistemas, portanto, não competem, mas cooperam, visto que há processos mentais que exigem o Sistema 1 e outros que necessitam do Sistema 2. Todavia, ocorre que, por conta do esforço exigido pelo Sistema 2 e da velocidade e eficácia do Sistema 1, é possível que, mesmo diante de questões complexas, aquele

¹³⁴ WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., p. 25.

¹³⁵ KAHNEMAN, Daniel, op. cit., p. 127; NUNES, Dierle; LUD, Natanael; PEDRON, Flávio. Desconfiando da imparcialidade dos sujeitos processuais, p. 50.

¹³⁶ NUNES, Dierle; LUD, Natanael, op. cit., p. 50.

¹³⁷ WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., pp. 28-29.

¹³⁸ WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., p. 35.

venha apenas a endossar generalizações deste. Quando ocorre tal inversão, os resultados podem ser desastrosos, pois parte-se de simplificações exageradas que levam a decisões subótimas. As heurísticas e vieses ocorrem, portanto, quando se utiliza o Sistema 1, enquanto crê-se estar utilizando o Sistema 2.¹³⁹

Isso posto, torna-se evidente como tais falhas cognitivas podem interferir no âmbito judicial. Especialmente em uma situação de excesso de demandas, juízes e demais sujeitos processuais não têm condições e tampouco incentivos de analisar detalhadamente todas as peculiaridades dos casos que chegam até eles, de forma que criam “atalhos mentais” para decidir. Tais atalhos ou heurísticas configuram “mecanismos de decisão pré-pronta” e servem não somente para reduzir suas cargas de trabalho mental, mas também para gerar sensações de coerência e conforto cognitivo.¹⁴⁰

Tomando-se como exemplo o processo penal, certas situações recorrentes (i.e. os “crimes de sempre”, como furto, roubo, tráfico, receptação etc.) levam a uma habituação de padrões que criam convicções antecipadas e fazem com que os demais detalhes sejam ignorados, pois tidos como supérfluos. A busca por coerência se traduz muitas vezes na utilização de brocardos, adágios, súmulas ou outros “mantras e lugares comuns”¹⁴¹, que exemplificam alguns dos mecanismos de conforto cognitivo à disposição dos julgadores. Além disso, experiências passadas podem “contaminar” situações presentes, tais como a credibilidade de um depoimento de uma testemunha já conhecida de outros casos. Em conjunto, tais fatores sinalizam a inexatidão da crença do ser humano como “uma ficção ultra racional”.¹⁴²

Estar ciente das limitações dos processos cognitivos humanos é, portanto, essencial à análise dos vieses em algoritmos. Uma inteligência artificial será tão boa quanto for o material por meio do qual ela é treinada, de forma que dados tendenciosos farão com que ela chegue a resultados igualmente ruins.¹⁴³ Mais que isso, a depender da forma como são implementados, algoritmos não apenas irão

¹³⁹ WOJCIECHOWSKI, Paola; ROSA, Alexandre, op. cit., p. 36; NUNES, Dierle; LUD, Natanael, op. cit., p. 52.

¹⁴⁰ ROSA, Alexandre Morais da. Guia do Processo Penal Conforme a Teoria dos Jogos, pp. 145-146.

¹⁴¹ ROSA, Alexandre Morais da. Guia do Processo Penal Conforme a Teoria dos Jogos, pp. 762-763.

¹⁴² ROSA, Alexandre Morais da. Guia do Processo Penal Conforme a Teoria dos Jogos, pp. 162-163.

¹⁴³ NUNES, Dierle; LUD, Natanael, op. cit., p. 148.

reproduzir o comportamento decisório humano, mas desenvolver seus próprios vieses e, inclusive, acentuar certas distorções.

3.2.2 Vieses em máquinas

Quando pessoas que não estudaram matemática veem notações, elas se sentem intimidadas, assustadas. Há aí um certo tipo de autoridade, algum tipo de objetividade, de verdade científica, que eles não estão autorizados a questionar, pois não são *experts*. Essa "autoridade do inescrutável" é traduzida também para algoritmos.¹⁴⁴

Até o momento, abordou-se a questão da implementação do aprendizado de máquina em atividades legais a partir dos pressupostos e desafios relativos ao seu bom funcionamento. Esta seção, contudo, tratará daqueles modelos que deram errado ou, mais especificamente, das condições que podem vir a tornar um modelo em uma "arma de destruição matemática".¹⁴⁵

Um modelo, conforme explica Cathy O'Neil, pode ser entendido como "uma representação abstrata de algum processo (...), que, independentemente de ser um programa de computador ou estar em nossas cabeças, utiliza informações que já sabemos para prever respostas em variadas situações".¹⁴⁶ Essencialmente, o que faz um modelo é utilizar dados preexistentes para descrever certas regularidades, que podem vir a ser úteis para processos de tomada de decisão futuros, à medida que tais padrões podem ser aplicados em novas situações. Dessa forma, todo modelo conta com *inputs* (dados de entrada), *outputs* (dados de saída ou predições) e uma definição de sucesso. Um modelo dinâmico, ao seu turno, é aquele que recebe constantemente novas informações e utiliza-as para aprimorar suas predições, com base em sua definição de sucesso.

Todavia, nenhum modelo é capaz de "captar toda a complexidade do mundo real ou as nuances da comunicação humana".¹⁴⁷ Seu objetivo, na verdade, é ser

¹⁴⁴ O'Neil Cathy. Weapons of Math Destruction: Vídeo-conferência. Personal Democracy Forum 2015. Disponível em: <https://www.youtube.com/watch?v=gdCJYsKIX_Y>. Acesso em: 16 nov. 2019. Tradução livre.

¹⁴⁵ Cf. O'NEIL, Cathy. Weapons of math destruction. New York: Broadway Books, 2016.

¹⁴⁶ O'NEIL, Cathy, op cit., p. 25.

¹⁴⁷ O'NEIL, Cathy, op cit., p. 26.

uma simplificação, que permita separar aspectos relevantes e irrelevantes de um determinado processo para se atingir um determinado fim. Assim, um sistema de mapas de um GPS para veículos terrestres deve possuir informações detalhadas sobre estradas, pontes e túneis, ao passo que não precisa levar em conta, por exemplo, o formato de prédios, pessoas ou temperatura. Já um *software* que guia aviões deve, dentre outras medições, saber a velocidade do vento e a temperatura local, mas não necessita de informações sobre pontes ou estradas.

Dessa forma, ao se criar um modelo, escolhem-se quais aspectos da realidade ele deve levar em conta e os demais para os quais isso não é preciso. As características deixadas de lado (“pontos cegos”) por modelos revelam as opiniões e prioridades de seus criadores, que são frutos de suas ideologias e ambições.¹⁴⁸ A isso soma-se o fato de que quanto mais complexo for o processo que o modelo busca descrever, mais variáveis ele deverá levar em conta, o que significa que exclusões arbitrárias podem resultar em simplificações errôneas.

É geralmente por “sobresimplificar” processos complexos que um modelo começa a apresentar anomalias, o que já revela um dos grandes desafios de buscar-se implementá-los no âmbito judicial, uma vez que este lida com os mais variados aspectos da vida humana.

Ainda assim, “se um modelo funciona ou não, também é uma questão de opinião”¹⁴⁹, pois sua definição de sucesso é arbitrária. Ocorre que um modelo pode estar funcionando bem aos olhos de seus programadores, mas não sob a perspectiva das pessoas que por ele são afetadas. Se uma companhia de seguros utiliza um algoritmo para cobrar mais caro por apólices para pessoas de um determinado grupo, o modelo estará funcionando bem aos olhos de seus administradores, os quais provavelmente estarão aumentando seus lucros, mas talvez não para aqueles que terão de pagar a mais pelos seus serviços.¹⁵⁰

Em modelos “inofensivos”, tais como aqueles utilizados em previsões desportivas, geralmente não há muito que se discutir em relação ao seu sucesso ou fracasso, uma vez que seu objetivo é bastante claro (e.g. ganhar jogos). O mesmo ocorre com um carro autônomo, que terá cumprido com êxito seu objetivo se tiver se

¹⁴⁸ O'NEIL, Cathy, op cit., pp. 26-27.

¹⁴⁹ O'NEIL, Cathy, op cit., p. 27.

¹⁵⁰ O'NEIL, Cathy, op cit., p. 28.

deslocado entre dois pontos sem causar acidentes. Todavia, especialmente quando modelos são aplicados em atividades com grande impacto social, muitas vezes em prejuízo das pessoas por eles afetadas, haverá debates sobre o que define seu bom funcionamento.

Fica evidente, destarte, que projetar modelos é uma tarefa permeada por subjetividade, ainda que seu grau possa variar de acordo com o tipo de processo a ser modelado. Em todo caso, para que um modelo tenha potencial de causar grande prejuízo para um grupo de pessoas, é preciso uma conjunção de fatores. Para Cathy O'Neil, um modelo se torna uma “arma de destruição matemática” quando ele é (i) opaco, (ii) possui a capacidade de crescer exponencialmente e (iii) é projetado para operar prejudicar aqueles a ele sujeitos.¹⁵¹

Especialmente quando utilizados pela Administração pública, modelos têm um grande potencial de se tornarem danosos. Afinal, por pressuposto, seus usos serão estendidos a um grande número de pessoas e, junto disso, modelos apresentam uma “tendência à inescrutabilidade”. Dessa forma, os dois primeiros requisitos estarão, em boa parte dos casos, automaticamente preenchidos ou em sua iminência. O terceiro requisito se torna, portanto, fundamental, uma vez que a forma como os modelos são projetados e os propósitos para os quais eles serão utilizados impactarão diretamente no seu potencial lesivo.

Não por coincidência, veio do judiciário (estadunidense) o primeiro caso utilizado por O'Neil para exemplificar esse tipo de modelo.¹⁵² Nos Estados Unidos, em um contexto de sobrerrepresentação populacional de negros em cadeias, aos quais são aplicadas penas, em média 20% maiores que a homens brancos condenados por crimes similares, vinte e quatro estados aderiram a algoritmos que prometem um sistema mais imparcial e livre de vieses, para tornar as sentenças judiciais mais homogêneas e previsíveis. Ademais, prender pessoas injustamente por mais tempo representa maiores gastos, de modo que tais algoritmos podem contribuir para economizar recursos públicos.¹⁵³

O mais famoso desses modelos, o LSI-R, inventado em 1995, conta com um longo questionário que os condenados devem responder, que será utilizado para

¹⁵¹ O'NEIL, Cathy, op cit., pp. 32-33.

¹⁵² Cf. O'NEIL, Cathy, op cit., Chapter 1 - Bomb parts: What is a Model?.

¹⁵³ O'NEIL, Cathy, op cit., pp. 29-30.

“prever” a probabilidade de que um condenado venha a reincidir, classificando tal risco em alto, médio ou baixo. Dentre as perguntas que devem ser respondidas pelos réus figuram: “quantas condenações anteriores você teve?” ou “qual a influência de drogas ou álcool na prática do crime?”. O questionário também traz questões relativas a circunstâncias de nascimento e formação do condenado, inclusive perguntas sobre o histórico criminal de membros de sua família e amigos.

Mesmo diminuído o fator aleatório decorrente de decisões humanas, alguém pode questionar se os vieses foram eliminados ou apenas camuflados. Ainda que não contenha perguntas que digam respeito expressamente à raça, não é difícil imaginar como um modelo que leva em conta tais variáveis apresentará um viés racial.

É evidente que pessoas vindas de bairros de baixa renda (mais policiados) terão maior probabilidade de já terem tido prévios contatos com a polícia, se comparados com pessoas de classe média ou alta. Igualmente, será mais provável que alguém de sua família contenha algum histórico criminal ou problemas relacionados a álcool e drogas. Nesse cenário, o modelo sequer precisa analisar uma resposta relativa à raça para se tornar racista, pois tal informação é desnecessária quando parte significativa das demais já aponta para questões racialmente sensíveis.

Em estados como Colorado e Idaho, o *score* do preso é utilizado por juízes para fundamentar suas sentenças. Todavia, tanto nos Estados Unidos quanto no Brasil, se um juiz fundamenta sua sentença com base no histórico criminal de um familiar do condenado, sua decisão seria imediatamente objeto de impugnação, pois, em ambos os países, é ilegal (e inconstitucional) julgar alguém por algo que ele não fez.¹⁵⁴

Acontece que, quando embutidas em um algoritmo, além de se revestirem de autoridade científica, tais opiniões passam geralmente despercebidas, pois não são inteligíveis para a maior parte das pessoas. Todavia, elas continuam (e sempre continuarão) a ser, simplesmente, opiniões, mas que agora estão perpetuadas e disfarçadas. Em um dado momento, alguém decidiu que eram relevantes para um

¹⁵⁴ Cf. BRASIL. Constituição (1988), art. 5º, XLV: “nenhuma pena passará da pessoa do condenado, podendo a obrigação de reparar o dano e a decretação do perdimento de bens ser, nos termos da lei, estendidas aos sucessores e contra eles executadas, até o limite do valor do patrimônio transferido”.

algoritmo que calcula a probabilidade de reincidência de um indivíduo informações relativas ao seu grau de escolaridade ou ao histórico criminal de seus pais. Nenhuma dessas conclusões é isenta de críticas e tampouco são “verdades científicas”.

Poder-se-ia argumentar que não se sabe ao certo (e realmente não se sabe) qual o peso que o algoritmo dá a essas questões, até porque, caso se soubesse, os condenados poderiam simplesmente alterar suas respostas para diminuir seus *scores*. Contudo, qualquer que seja tal peso, ele será injusto (e ilegal).¹⁵⁵ Ainda que isso represente uma possível melhora com relação ao preconceito e aleatoriedade de juízes humanos, tais modelos continuam sendo injustos e isso não deve ser esquecido.

Ainda que se coloque momentaneamente de lado o critério da justiça, tais modelos apresentam mais um grave problema. Suas opiniões embutidas acabam por recriar as pressuposições que as sustentam, o que caracteriza um “ciclo de retroalimentação vicioso”.¹⁵⁶ Uma pessoa vinda de uma região socioeconomicamente vulnerável possuirá maiores chances de ser taxada com um alto grau de possibilidade de reincidência, o que lhe fará ser condenada por mais anos. Ao ficar mais tempo afastada do convívio social e em contato com outros prisioneiros, ela terá maiores dificuldades de voltar a se inserir na sociedade e conseguir um emprego, por exemplo. Dessa forma, caso cometesse outro crime e voltasse à cadeia, o algoritmo entenderia que realizou uma predição correta e reforçaria seu próprio entendimento.

Esse *loop* vicioso, além de contribuir, no caso em análise, para que pessoas sejam punidas por conta de fatos que ainda não ocorreram, piora ainda mais o problema, uma vez que recria e fortalece o ambiente que justifica suas premissas. Esse aspecto é frequentemente presente em “armas de destruição matemática” e está intimamente relacionado ao aspecto destrutivo de tais modelos.

Como, então, evitar que algoritmos se convertam em uma ameaça a toda uma sociedade? Como dito, tais modelos são preconcebidos para serem escaláveis a toda a sociedade, o que torna incontornável a questão de seu crescimento

¹⁵⁵ O'NEIL, Cathy, op cit., p. 30.

¹⁵⁶ O'NEIL, Cathy, op cit., p. 30.

exponencial. Com relação à sua opacidade, em um Estado democrático e de direito, ela será sempre um desafio. O simples fato de um modelo ser escrito através de uma notação matemática e/ou por meio de uma linguagem computacional já o torna inacessível à maior parte das pessoas. Esse problema se torna ainda mais pertinente, quando se trata de modelos que utilizam aprendizado de máquina, cujos detalhes de seu modo de funcionamento não são conhecidos nem mesmo por seus programadores.

Especialmente quando utilizados pela Administração pública, algoritmos devem seguir certos padrões mínimos de transparência e prestação de contas, bem como os princípios de direito que orientam o agir administrativo. Ainda assim, isso não significa que modelos não possam ser tornados compatíveis com governos democráticos.

Divide-se em ao menos duas etapas o processo de prestação de contas dos modelos, ainda que se queira preservar a integralidade de seus códigos propriamente ditos. Em um momento pré-elaboração dos códigos, é fundamental que a sociedade em geral tenha ciência de quais aspectos da realidade serão levados em conta pelo modelo, o que é percebido nos dados que serão utilizados como *inputs*. Dessa forma, é possível, por exemplo, contestar os motivos que levaram a certos dados serem considerados pertinentes e as razões pelas quais outros não foram sequer considerados.

Após escritos tais códigos, pode-se fiscalizá-los por meio da auditoria de seus resultados, momento em que são detectados eventuais vieses ou distorções. Assim, uma vez implementado um modelo e verificado que ele penaliza um certo segmento social desproporcionalmente aos demais, sem razão justificada, pode-se pleitear sua alteração ou seu desativamento. É claro que o ideal seria uma atuação preventiva e não paliativa, mas isso não diminui a importância de sempre se estar verificando os *outputs* obtidos. Para tanto, mais uma vez, é necessário que tais resultados estejam acessíveis ao público, especialmente àqueles que são diretamente afetados.

Por fim, e talvez mais importante, é necessário que a própria ideia de concepção do algoritmo provenha de uma construção conjunta entre todas as partes interessadas, a saber, gestores, público em geral e desenvolvedores. A participação de diversos setores da sociedade na elaboração de algoritmos é essencial para que

suas definições de sucesso e as tarefas nas quais eles são empregados sejam estabelecidas da forma mais benéfica ao maior número de pessoas ou, quando isso não seja possível, sejam implementados de forma a mitigar ao máximo seus malefícios.

Dessa forma, diminui-se a chance de que tais algoritmos se tornem máquinas projetadas para arruinar vida de pessoas ou que entrem em ciclos de retroalimentação viciosos. Deve-se observar que um mesmo algoritmo, treinado com os mesmos dados, pode ser mais ou menos danoso, a depender do uso que se faz dele. O citado LSI-R, ainda que o questionário que o fundamenta seja repleto de questões controversas, tem seu potencial danoso reduzido em estados que o utilizam não para fundamentar sentenças, mas para selecionar presidiários que serão incluídos em programas de prevenção de reincidência enquanto estão cumprindo suas penas.

Dessa forma, é imprescindível que se questione acerca do fim para o qual um modelo será utilizado. Mais uma vez, a resposta só pode vir de uma construção conjunta entre todos os interessados. Isso nada mais é do que um dos pressupostos de um governo democrático e deveria ser adotado como padrão sempre que decisões que afetam um grande número de pessoas são tomadas.

3.3 Formas de se utilizar o aprendizado de máquina no judiciário

A partir das análises sobre linguagem, direito e algoritmos expostas nos tópicos anteriores, podem-se elencar alguns dos tipos de uso do aprendizado de máquina em atividades relacionadas à jurisdição e suas respectivas vantagens e desvantagens. Antes de mais nada, deve-se observar que a complexidade do fenômeno linguístico representa por si só um entrave ao processamento de linguagem natural, sobretudo em atividades jurídicas, que possuem diversas especificidades no que diz respeito a formas de se definir conceitos e argumentar.¹⁵⁷

É verdade que a utilização de algoritmos contribui para minimizar fatores externos aleatórios tipicamente humanos, tais como cansaço e instabilidade emocional, mas eles também estão sujeitos a vieses estruturais decorrentes do

¹⁵⁷ Cf. seção 1.2 deste trabalho.

sistema jurídico, da forma como eles são treinados e de sua própria programação.¹⁵⁸ Como visto, ainda que algoritmos como o Word2vec sejam capazes de assimilar, ao menos em parte, o contexto textual de palavras, não se pode afirmar que eles compreendam conceitos da forma como humanos fazem. Sua “compreensão” limita-se a associar uma palavra a outras que geralmente a acompanham e, ainda que se possa chegar a bons resultados através desse método, isso não é o suficiente para dar conta de todas as formas de uso da linguagem, que, assim como o direito, configura um fenômeno social complexo.

Disso decorre a dificuldade de se quantificar em números certos conceitos, que, na verdade, não são computáveis, pois decorrem de um certo “jogo” que não pode ter suas regras previamente definidas. Uma vez que “computadores, a despeito de todos os seus avanços em linguagem e lógica, ainda têm muitos problemas com conceitos”¹⁵⁹, eles acabam tendo que utilizar dados “aproximados” para se alcançar uma possível definição termos abstratos.¹⁶⁰

Todavia, ao assim proceder, há o risco de perpetuação de um *status quo* relativo a uma certa ordem social¹⁶¹, à medida que se pereniza uma certa concepção de mundo, dentre diversas possíveis. Por conta disso, nem sempre será possível conciliar eficiência com equidade, de modo que será necessário optar, em algumas situações, entre uma ou outra. Nesses casos, em um ambiente democrático, caberá à sociedade definir o que se espera do uso de algoritmos em atividades que impactam um grande número de pessoas. Sua atuação será fundamental para definir as balizas que guiarão tais práticas e deverá vir dela a palavra final sobre o que se está disposto a sacrificar em benefício da eficiência.

3.3.1 A questão da corrobóica

Em todo caso, encaminha-se para um cenário em que a não utilização de algoritmos será um sinônimo de obsolescência e o poder judiciário se mostra um dos

¹⁵⁸ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, p. 44.

¹⁵⁹ Tradução livre: “And computers, for all of their advances in language and logic, still struggle mightily with concepts.” em O’NEIL, Cathy, op cit., p. 82.

¹⁶⁰ O’NEIL, Cathy, op cit., p. 82.

¹⁶¹ O’NEIL, Cathy, op cit., p. 79.

terrenos mais propícios para a implementação do aprendizado de máquina. Isso se deve ao fato de que nele concorrem fatores como: i) grande concentração de dados potencialmente tratáveis ii) grande demanda por agilidade na prestação do serviço; iii) disponibilidade de orçamento para implementar soluções inovadoras e; iv) escassez de recursos humanos para cumprir com a carga de trabalho demandada.

Antes de mais nada, deve-se levar em conta que as limitações atuais do aprendizado de máquina, inseridas dentro da lógica de uma inteligência artificial de propósito limitado, somente permitem que juízes sejam substituídos em situações específicas, uma vez que tais profissionais executam tarefas de diferentes naturezas, excedendo, portanto, o escopo de atuação daquele tipo de inteligência.¹⁶² Além disso, não se pode esperar que máquinas atuem de forma perfeita, bastando que elas contem com precisão e transparência coerentes com o tipo de atividade a ser desempenhada e que isso ocorra em um grau próximo de processos equivalentes executados por seres humanos.

Dessa forma, dado que, ao menos por hora, juízes humanos não poderão ser completamente eliminados, as questões relativas ao uso do aprendizado de máquina na administração da justiça dizem respeito principalmente à coexistência entre humanos e máquinas em um mesmo ambiente e as consequências daí decorrentes, o que define a *corrobótica*. Nesse sentido, torna-se relevante a forma como os algoritmos influenciarão o processo decisório levado a cabo por humanos, bem como saber se isso significa alguma renúncia de poder por parte do ser humano.

Todavia, a depender da abordagem escolhida, é possível contornar a questão da *corrobótica*. Em casos excepcionais, será buscado eliminar o fator humano da elaboração de decisões, ainda que estas tenham que ser revistas posteriormente por juízes de carne e osso (o que não remove por completo, portanto, o *homo sapiens* da equação). Isso ocorrerá nos casos em que “juízes-robô” irão decidir litígios do início ao fim e humanos se tornam uma espécie de instância recursal. Ainda assim tal tipo de uso será restrito a casos de baixa complexidade e que comportam pouco ou nenhum poder discricionário por parte dos juízes (casos que

¹⁶² BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, p. 46.

não fogem, portanto, da aplicação padrão da norma e que dificilmente geram debates).

Nesse caso, o problema não diz mais respeito à corrobótica, mas a novas questões. Haverá uma “pressão” para que juízes humanos ratifiquem decisões das máquinas? O que ocorre em caso de divergência? Além disso, um caso dito simples para uma pessoa pode ser complexo para outras. Quem decidirá o tipo de litígio que poderá ser objeto de uma decisão automatizada?

Por conseqüente, a depender da abordagem utilizada, i.e. máquinas auxiliando humanos no processo decisório ou tomando decisões sozinhas, evita-se o problema da corrobótica ou se adere ao mesmo. Ambas as abordagens, de todo modo, devem lidar com o fato de que algoritmos de aprendizado de máquina gozam de certa autonomia, aqui entendida como a capacidade de ditar (ao menos em parte) suas próprias regras.

Contudo, na primeira abordagem, torna-se especialmente importante a questão da transparência da comunicação entre o algoritmo e o humano. É certo que a transparência só ocorre ao custo da performance, pois significa forçar uma inteligência artificial a trabalhar com parâmetros humanos, os quais nem sempre são os mais eficientes.¹⁶³ Conforme explanado, as inferências causais que o algoritmo estabelece nem sempre são as mesmas que ocorrem no mundo real e, portanto, podem ser incompreensíveis a humanos.

Todavia, nem mesmo um juiz humano é totalmente transparente, se por “transparência” se entende compreender cada um dos impulsos neurológicos que levam uma certa pessoa a tomar uma certa decisão. Deve-se, dessa forma, distinguir entre transparência no que tange aos detalhes do funcionamento de uma máquina daquilo que diz respeito à capacidade de se explicar seus resultados, ou seja, a decisão gerada pela máquina, o que não é impossível de ser feito.

Já no caso da abordagem que evita o problema da corrobótica, os juízes-robô não necessariamente necessitam ser transparentes, uma vez que a instância humana revisora cumprirá esse papel. Todavia, não se pode ignorar que, tanto mais seja possível que robôs “expliquem” suas decisões, maior será a legitimidade das

¹⁶³ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, p. 49.

mesmas. Transparência, nessa situação, significa que os algoritmos sejam capazes de expor termos ou frases que tiveram maior peso na classificação do caso, bem como quais leis ou precedentes foram aplicados em sua resolução. Mais uma vez, não pode se esperar aqui uma transparência no nível das minúcias do funcionamento, mas somente no que diz respeito a tornar algumas de suas razões humanamente inteligíveis.

3.3.2 Três tipos de uso do aprendizado de máquina no direito

Dessa forma, verificam-se diversos usos potenciais do aprendizado de máquina no âmbito judicial, alguns dos quais já foram ou estão em vias de ser implementados. Dentre as possíveis formas de utilização, podem-se elencar três grandes grupos (ou tipos de uso), que concentram a maior parte dessas abordagens. Para se definir tais grupos, foram escolhidos alguns critérios relativos a questões normativas e técnicas, a saber, (i) *grau de intervenção humana*, (ii) *interferência do algoritmo no processo decisório*, (iii) *complexidade do algoritmo envolvido* e (iv) *transparência da decisão*.

Antes de mais nada, deve-se notar que alguns dos critérios não são absolutos, mas apenas comparações com seus equivalentes em processos decisórios executados por seres humanos. Assim, quando se fala em um “alto grau de transparência”, quer-se dizer que a transparência do processo decisório (iv) ocorre a um nível semelhante ao de uma decisão elaborada por seres humanos sem a intervenção de algoritmos. O mesmo ocorre com relação ao grau de intervenção humana (i).

Um primeiro tipo, doravante denominado *classificador*, tem por função primordial encontrar materiais úteis para que humanos fundamentem suas decisões. Tais materiais compreendem, por exemplo, dispositivos normativos, precedentes judiciais e modelos de documentos que servirão de base para direcionar o pronunciamento judicial.

Além disso, tais algoritmos podem ser utilizados para poupar recursos financeiros e humanos ao identificar processos em tramitação como pertinentes a um determinado tema de Repercussão Geral (RG), Recurso Repetitivo (RR),

Incidente de Resolução de Demandas Repetitivas (IRDR) ou Incidente de Assunção de Competência (IAC), em conjunto aqui denominados “sistema de vinculação de precedentes”. Dessa forma, pode-se sobrestar sua tramitação até que o respectivo tema seja estatuído e, com isso, evitar que os processos já julgados tenham de ser revistos por estarem em sentido contrário ao decidido por tribunais superiores.

Para esse tipo de uso, a complexidade dos algoritmos envolvidos é relativamente baixa. Desde que haja uma quantidade razoável de dados tratáveis, pode-se, sem maiores dificuldades, treinar máquinas para identificar peças processuais (petições iniciais, apelações, sentenças, etc) e estimar sua relevância para outro caso ou sua pertinência a algum tema relativo ao sistema de vinculação de precedentes. Dessa forma, podem-se identificar, em um acervo de milhões de documentos, aqueles mais relevantes para um determinado litígio.

Dado que humanos terão de elaborar integralmente os documentos, sua intervenção no processo decisório será máxima. Ainda assim, isso não significa que não haverá interferência da máquina, pois esta ainda pode induzir o entendimento dos humanos à medida que “filtra” a informação à qual eles terão acesso, o que é consequência das considerações expostas na seção anterior.

Por exemplo, quando se treina algoritmos para encontrar processos vinculados a algum tema de precedentes, na verdade está-se sedimentalizando um certo entendimento acerca daquele determinado tema. Assim, quando um humano julgar um novo processo, receberá como sugestão do algoritmo um outro processo dito “semelhante”. Já houve aí uma interferência, pois, dentre os eventuais diversos entendimentos acerca daquele caso, foi escolhido um deles.

É claro que o juiz pode discordar da classificação, mas a tendência será de que a sugestão do algoritmo se transforme em novas decisões judiciais, em um processo recíproco de seleção informação e sua transformação em expertise legal.

¹⁶⁴ Uma vez publicada, tal decisão passará a compor o conjunto de processos relativos àquele tema e passará a ser sugerida em novos casos, retroalimentando um entendimento prévio.

¹⁶⁴ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, pp. 51-52.

Por fim, a transparência desse processo ocorrerá sempre em nível próximo ao humano. Da mesma forma que um juiz pode pedir ao seus assessores que eles procurem manualmente precedentes relativos a um determinado caso, ele pode se fiar em um algoritmo para que o faça. Em ambos os casos, será necessário que ele fundamente sua decisão final da mesma forma como tradicionalmente faz. Ainda assim, é importante que o algoritmo forneça meios para que suas decisões sejam interpretáveis, uma vez que a pesquisa legal e o processo decisório são atividades interdependentes.¹⁶⁵

Como exemplo desse modo de se utilizar o aprendizado de máquina, tem-se o projeto Victor, do Supremo Tribunal Federal, que tem por escopo trazer maior agilidade na tramitação de processos na Corte Superior ao sugerir automaticamente vinculações de processos novos a temas de Repercussão Geral (RG).¹⁶⁶ Pode-se facilmente imaginar, em um futuro próximo, que tribunais e outros órgãos vinculados à administração da justiça de todo o país implementem técnicas semelhantes ou que utilizem tecnologias similares para resgatar suas próprias decisões, na tentativa de homogeneizar sua jurisprudência.

Uma segunda forma de utilização do aprendizado de máquina, aqui apelidada de *relatora*, diz respeito a extrair e condensar informações relevantes de um certo documento, o que pode ser utilizado para diferentes fins. Para tanto, ela deve ser capaz não apenas de encontrar documentos similares, mas ir mais a fundo em sua estrutura, diferenciando, em cada peça processual, aquilo que se refere à descrição de fatos, textos legais, jurisprudências colacionadas e estruturas argumentativas.

Nesse sentido, é necessário que o algoritmo possua habilidades em, ao menos, mineração de textos, expansão de conceitos e extração de relações¹⁶⁷, isso porque se torna necessário ser capaz de encontrar informações-chave que resumem um documento, sintetizar argumentos das partes e identificar relações semânticas e sintáticas entre os termos.

¹⁶⁵ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, pp. 51-52.

¹⁶⁶ SUPREMO TRIBUNAL FEDERAL. Inteligência artificial vai agilizar a tramitação de processos no STF. Notícias STF. Brasília. 30 maio 2018. Disponível em: <<http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=380038>>. Acesso em: 01 nov. 2018.

¹⁶⁷ Cf. item 1.1.1 deste trabalho

Utilizado dessa maneira, um algoritmo “parecerista” ainda estaria apenas auxiliando um juiz na tarefa de fabricar uma decisão, de forma que persiste a questão da corrobóptica. Suas considerações podem ser aceitas ou recusadas, mas, frise-se, o juiz, como tendência, concordará com a máquina¹⁶⁸, seja por conveniência, seja pelo fato de que uma decisão amparada por um algoritmo terá menores chances de ser revista.

Nesse caso, a transparência se mantém a níveis próximos dos casos decididos sem o auxílio de algoritmos, pois o juiz continua tendo que dar os “toques finais” no documento (ou refazê-lo por completo). Ainda assim, se comparado ao uso anterior, é mais fácil saber a real extensão da utilização da inteligência artificial.¹⁶⁹

Contudo, uma vez aceitas as sugestões do algoritmo, a intervenção humana ocorre de forma significativamente mais baixa que no primeiro exemplo (classificador), dado que em muitos casos o juiz terá apenas de assinar o documento, limitando sua intervenção à revisão de alguns pontos e à averiguação de informações. Com o passar do tempo, a tendência é que a concordância se torne a regra, à medida que o modelo aprimora sua acurácia e são desenvolvidas novas técnicas de NLP. Em tal cenário, a interferência do algoritmo no processo decisório será maior que aquela da abordagem anterior e o juiz pode se tornar uma espécie de “canal de entrega” de decisões geradas por robôs.¹⁷⁰

Ademais, como dito, um algoritmo dessa natureza será mais complexo que o anterior, pois necessita de diversas habilidades que excedem a mera classificação de documentos. Todavia, sua maior complexidade significa uma maior versatilidade, que permite empregar algoritmos “relatores” em diferentes funções. A primeira diz respeito a elaborar decisões “pré-fabricadas” para juízes: a máquina indica ao juiz as páginas em que se encontram as peças processuais, elenca os argumentos trazidos por cada uma das partes e, eventualmente, sugere uma decisão para o caso.

¹⁶⁸ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, pp. 54-55.

¹⁶⁹ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, p. 54.

¹⁷⁰ BUOCZ, Thomas Julius. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary, p. 55.

Uma segunda possibilidade de aplicação desses algoritmos pode ocorrer de modo próximo à atuação de um juiz leigo. Por exemplo, em tribunais de pequenas causas, após ambas as partes terem tido a oportunidade de se manifestar, não sendo necessária a produção de novas provas e saneado o processo, um algoritmo sugere um encaminhamento ao caso, baseado em decisões passadas daquela corte. Se as partes concordarem, o acordo é encaminhado a um juiz humano para ratificá-lo, caso contrário, aguarda-se o julgamento humano. Deve-se observar que, dados os atuais limites tecnológicos, tanto no primeiro caso, quanto no segundo, os processos analisados pelos algoritmos deverão ser de baixa complexidade, limitando-se a questões como a revisão de multas, execuções fiscais, acidentes de trânsito etc.

Em outra forma de atuação, os robôs pareceristas podem atuar na predição de decisões judiciais (i.e. jurimetria). Uma vez que são capazes de diferenciar diferentes estruturas textuais, é possível treiná-los para analisar tipos de argumentação legal, descrição dos fatos, bem como provas acostadas aos autos para calcular-se quão relevantes eles foram para a procedência ou improcedência de ações passadas. Assim, em um novo caso, podem-se estimar as chances de se sair vitoriosa uma determinada linha argumentativa.

Nessa linha, pesquisadores conseguiram prever com acurácia média de 79% decisões do Tribunal Europeu de Direitos Humanos - TEDH.¹⁷¹ O experimento consistiu na elaboração de um modelo de classificação binária que, tomando como *inputs* exclusivamente dados textuais, dizia se houve ou não violação de alguns dos artigos da Convenção Europeia dos Direitos Humanos - CEDH.

Levando em conta a estrutura dos julgados da corte, os pesquisadores extraíram de decisões prévias informações relativas a questões procedimentais, fatos (circunstâncias do caso) e à letra da lei e utilizaram-nas na predição da parte dispositiva dos mesmos casos.¹⁷² Com isso, concluíram que não apenas é possível estabelecer uma correlação entre dados textuais de um caso e sua decisão, mas

¹⁷¹ Cf. ALETRAS, Nikolaos et al. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective.

¹⁷² ALETRAS, Nikolaos et al, op. cit., pp. 6-8.

também que as circunstâncias fáticas tinham mais peso na predição que os argumentos legais.¹⁷³

O terceiro e último tipo uso do aprendizado de máquina no judiciário, a saber, o robô-juiz, apresenta características muito próximas do modelo anterior (parecerista), no que diz respeito às suas funcionalidades. Sua diferença reside principalmente no tratamento que se dá ao resultado gerado pelo algoritmo, que, seja por ter atingido uma acurácia bastante alta, seja pelo processo ser de baixa complexidade, será tido como a própria decisão judicial.

Dessa forma, ocorre um ato completamente automatizado, através do qual se decide um processo judicial. No caso das partes discordarem do conteúdo decisório, apela-se à instância humana revisora, que, por sua vez, poderá manter ou reformar a decisão artificialmente gerada e o processo segue normalmente seu curso, a partir daí.

Por um lado, pode-se considerar que a interferência algorítmica no processo decisório é eliminada por completo, ao mesmo tempo que não há mais sentido falar-se em grau de intervenção humana, uma vez que se separa o componente humano da máquina. Dessa forma, ao contrário do que acontece com o “robô parecerista”, evita-se uma certa perda de poder decisório por parte do humano, que deriva da dificuldade de se discordar de uma certa “autoridade” que os algoritmos possuem por conta de suas pressupostas precisão e neutralidade.

Além disso, elimina-se o problema de justificação da decisão, pois o algoritmo apenas gerará um resultado, cujas explicações ficarão a cargo do juiz de carne e osso, se for o caso de inconformismo de uma das partes.¹⁷⁴ Nessa situação, a transparência da decisão seria máxima, pois, uma vez passada pelo crivo humano, suas justificativas seriam elaboradas como de costume.

Sob outra perspectiva, a influência da máquina no processo decisório poderá ser entendida como bastante elevada. Será muito mais conveniente aos juízes humanos ratificarem a sentença artificial, haja vista que eles estarão possivelmente tratando de casos massificados e de baixa complexidade, além de que será simplesmente mais trabalhoso discordar e explicar a falha da inteligência artificial.

¹⁷³ ALETRAS, Nikolaos et al, op. cit., pp. 15-16.

¹⁷⁴ BUOCZ, Thomas Julius, op. cit., pp. 55-56;

Em segundo lugar, ainda que de forma extraoficial, outra tendência é que se estabeleça uma “jurisprudência algorítmica”, que sirva de fundamento para novas decisões automáticas, em um processo cíclico, e da qual discordar poderá resultar em abrir precedentes para que decisões sejam revistas em milhares de outros casos.

Além disso, não seria surpreendente se decisões automaticamente elaboradas tivessem índices altíssimos de recorribilidade, especialmente no caso de elas não conterem em si mesmas justificativas satisfatórias e compreensíveis para humanos. Ainda assim, tais índices dependeriam da matéria em questão e da logística processual em vigor. Em todo caso, seria necessária uma análise minuciosa do custo-benefício da implementação desse tipo de sistema e dos resultados práticos esperados.

Um segundo problema diz respeito a quais casos poderão ser objeto de decisões automáticas. Se há discordância entre o humano e a máquina, algum deles errou, ambos erraram ou o caso apresenta mais de uma solução. O que difere caso simples (*plain cases*) de casos complexos (*hard cases*) é se a aplicação da norma se dá de forma automática e sua solução se dá de maneira simples ou se ele configura uma daquelas situações em que há, ao menos aparentemente, mais de uma decisão juridicamente válida.¹⁷⁵

Dessa forma, em *hard cases* não será possível dizer qual dos dois errou, pelo menos enquanto não houver uma decisão de uma corte superior para definir a matéria. Ainda assim, nesses casos, vale mais a explicação das razões que levaram a uma determinada decisão que a precisão (ou acurácia) da decisão em si. Por conta disso, *Hard cases* tendem a ser incompatíveis com a aplicação do aprendizado de máquina.¹⁷⁶

¹⁷⁵ BUOCZ, Thomas Julius, op. cit., p. 56.

¹⁷⁶ BUOCZ, Thomas Julius, op. cit., p. 56.

Conclusão

Buscou-se, com este trabalho, fornecer as bases de uma discussão multidisciplinar sobre o uso de ferramentas de aprendizado de máquina em atividades relacionadas à administração da justiça. Para tanto, foram levados em conta os aspectos técnicos de tais ferramentas, assim como a própria natureza da atividade interpretativa da linguagem natural, tarefa esta que se mostra um grande desafio, tanto para humanos, quanto para máquinas.

Em primeiro lugar, pôde-se observar que o aprendizado de máquina apresenta características peculiares, se comparado a outros modelos estatísticos. Primeiramente, ele possui certa autonomia, pois prescinde de que humanos expliquem previamente como se dão as relações entre as variáveis que compõem o modelo, ao mesmo tempo que pode ajustar seus parâmetros (pesos e vieses) de maneira autônoma para melhor descrever os dados que tem à sua disposição. Conjuntamente, tais características permitem que técnicas de *machine learning* executem tarefas cujas minúcias são enigmáticas mesmo para humanos.

Todavia, seu melhor desempenho é acompanhado de um custo interpretativo. Não é possível descrever passo a passo como tais algoritmos chegam a uma determinada decisão, aspecto que lhes confere a alcunha de “caixa-preta”. Ainda assim, isso não significa que é impossível interpretar seus resultados, visto que, em vários casos, pode-se estimar quais variáveis tiveram maior peso para se chegar a um dado resultado.

No âmbito do direito, o aprendizado de máquina é utilizado principalmente para ensinar computadores a “ler” textos escritos por humanos, o que configura o processamento de linguagem natural, um dos subcampos da inteligência artificial. Máquinas, contudo, não interpretam textos da mesma forma que humanos. Sua forma de assimilar a linguagem natural consiste em criar representações matemáticas de palavras, através das quais é possível identificar relações semânticas e sintáticas entre termos e similaridades entre documentos.

Ainda assim, a leitura e a redação jurídicas compõem tarefas mais complexas que simplesmente extrair informações de textos longos ou formular frases

gramaticalmente bem escritas. O raciocínio legal (*legal reasoning*) requer, por exemplo, fundamentações detalhadas acerca dos motivos que fizeram um argumento (e não outro) ser acatado, bem como levar em conta a hierarquia de leis e decisões judiciais prévias.

Por conta disso, torna-se imprescindível que máquinas possuam habilidades relativas à mineração de textos, tais como extrair informações de textos longos, encontrar respostas a perguntas específicas, minerar de argumentos, expandir conceitos, dentre outras. Por meio do aprendizado de máquina, tornou-se possível extrair tais informações de modo mais eficiente e aplicá-las a grandes quantidades de documentos, através de modelos que funcionam conjuntamente com a atuação humana, o que leva ao paradigma da corrobótica ou da computação cognitiva.

Ainda assim, a discussão sobre o uso do ML em tribunais não se restringe somente aos seus aspectos técnicos. Defendeu-se que a forma como se pensa a linguagem impacta diretamente no modo de se conceber o fenômeno jurídico. Como exemplo, podem-se citar os casos de Hans Kelsen e Herbert Hart, que, a partir do pano de fundo das discussões, respectivamente, do Círculo de Viena e da filosofia da linguagem ordinária, propuseram diferentes considerações acerca do fenômeno jurídico.

Ainda assim, ambos os jusfilósofos chegam a um ponto comum, ainda que por caminhos diferentes, a saber, que há casos em que a imprecisão da linguagem cria uma margem de discricionariedade àqueles que aplicam as normas. A inafastabilidade da discricionariedade pode ser entendida como uma consequência dos pressupostos da filosofia analítica, dentro do qual Kelsen e Hart estão inseridos.

Uma vez que o fenômeno jurídico é encarado não como uma "ciência exata", mas como uma atividade sujeita à discricionariedade, surgem novos questionamentos sobre o uso do aprendizado de máquina e do processamento de linguagem natural. Em primeiro lugar, deve-se ter em mente que linguagens de programação operam em apenas dois níveis semióticos, i.e. sintático e semântico, enquanto linguagens naturais possuem, para além destes, a dimensão pragmática. Ainda assim, o âmbito pragmático da linguagem ordinária não é completamente inacessível a algoritmos, uma vez que estes podem ser treinados de forma a levar

em conta o contexto textual em que um termo ocorre, o que lhes confere um melhor desempenho, tanto linguístico, quanto computacional.

Em segundo lugar, a existência de certo grau de discricionariedade na interpretação de termos, e, conseqüentemente, de normas jurídicas, levanta indagações sobre o modo como isso ocorre em tribunais e de que forma isso interfere na elaboração de algoritmos. Nesse sentido, são valiosas as contribuições da psicologia cognitiva, que explica de que forma operam heurísticas e vieses em processos mentais decisórios. Grosso modo, pode-se dizer que o cérebro humano cria atalhos cognitivos, por meio dos quais é reduzido o esforço necessário para tomar decisões complexas, mecanismo bastante útil no dia a dia das pessoas. Todavia, em outras situações, tais como decisões judiciais, esse mecanismo pode levar à tomada de decisões subótimas, ao passo que muitas vezes ignora a quantidade e a qualidade das informações disponíveis.

Dado que os algoritmos do judiciário são treinados com base em decisões humanas, eles podem reproduzir ou, ainda, acentuar tais vieses. Isso se torna especialmente perigoso quando eles são vistos como ferramentas imparciais e revestidas de cientificidade. Para evitar que os mesmos se tornem "armas de destruição matemática", é necessário velar para que tais modelos sejam elaborados de forma transparente e conjunta com todos aqueles que serão afetados por suas decisões, bem como que seus resultados sejam cuidadosamente auditados.

Por fim, foram elencados três abordagens mais recorrentes do uso do aprendizado de máquina no judiciário. A primeira delas corresponde ao uso classificador, através do qual algoritmos auxiliam humanos em tarefas básicas, tais como encontrar e classificar processos. A abordagem do robô-parecerista consiste em condensar informações relevantes de um processo em um único documento, que eventualmente pode ser utilizado para sugerir decisões a um certo caso concreto. A terceira e última, robô-juiz, consiste em um passo adiante com relação à anterior, de forma que seus resultados são considerados vinculativos e elimina-se do processo decisório completamente o componente humano, que se torna uma instância revisora.

Referências

ASHLEY, Kevin D.. **Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age**. University Of Pittsburgh School Of Law: Cambridge University Press, 2017. (ISBN: 9781107171503).

ALETRAS, Nikolaos et al. **Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective**. Peerj Computer Science 2: <https://doi.org/10.7717/peerj-cs.93>, 2016.

BRASIL. **Constituição (1988)**. Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

BRASIL. CONSELHO NACIONAL DE JUSTIÇA. **Justiça em Números 2019: Relatório Analítico**. Brasília: Departamento de Pesquisas Judiciárias, 2019. Disponível em:

<https://www.cnj.jus.br/wp-content/uploads/conteudo/arquivo/2019/08/justica_em_numeros20190919.pdf>. Acesso em: 16 nov. 2019.

BELLONI, Massimo. **Neural Networks and Philosophy of Language**. 2019. Disponível em:

<<https://towardsdatascience.com/neural-networks-and-philosophy-of-language-31c34c0796da>>. Acesso em: 16 nov. 2019.

BUOCZ, Thomas Julius. **Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary**. Retskraft: Copenhagen Journal of Legal Studies. Copenhagen, p. 41-59. mar. 2018. Disponível em:

<<https://static1.squarespace.com/static/59db92336f4ca35190c650a5/t/5ad9da5f70a6adf9d3ee842c/1524226655876/Artificial+Intelligence+in+Court.pdf>>. Acesso em: 16 nov. 2019.

CABRERA, Julio. **Margens das Filosofias da Linguagem**. Brasília: Editora Universidade de Brasília, 2003. 320 p.

COGLIANESE, Cary and LEHR, David. **Regulating by Robot: Administrative Decision Making in the Machine-Learning Era**. Georgetown Law Journal, Vol. 105, p. 1147, jun. 2017; U of Penn, Inst for Law & Econ Research. Paper No. 17-8. Disponível em SSRN: <<https://ssrn.com/abstract=2928293>>. Acesso em 01 nov. 2018.

FIDALGO, António. **Semiótica: A Lógica da Comunicação**. 1998. 135 f. Universidade da Beira Interior, Covilhã, 1998.

GANDHI, Rohith. **Support Vector Machine — Introduction to Machine Learning Algorithms**. 2018. Disponível em: <<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>>. Acesso em: 16 nov. 2019.

GONTIJO, Lucas de Alvarenga. **Filosofia do Direito: metodologia jurídica, teoria da argumentação e guinada linguístico-pragmática**. Belo Horizonte: Arraes, 2011.

GERMANO, Felipe. **Computador vence humano em Go, jogo mais complexo que xadrez**. Super Interessante. São Paulo. 04 nov. 2016. Disponível em: <<https://super.abril.com.br/tecnologia/computador-vence-humano-em-go-jogo-mais-complexo-que-xadrez/>>. Acesso em: 01 nov. 2018.

GROSSFELD, Brett. **A simple way to understand machine learning vs deep learning**. 2017. <https://www.zendesk.com/blog/machine-learning-and-deep-learning/>. Disponível em: <<https://www.zendesk.com/blog/machine-learning-and-deep-learning/>>. Acesso em: 01 nov. 2018.

HART, Herbert L. A.. **O Conceito de Direito**. 3. ed. Lisboa: Fundação Calouste Gulbenkian, 2001. Tradução de A. Ribeiro Mendes.

KAHNEMAN, Daniel. **Rápido e devagar: duas formas de pensar**. Rio de Janeiro: Editora Objetiva, 2011. Tradução de Cássio de Arantes Leite.

KARANI, Dhruvil. **Introduction to Word Embedding and Word2Vec**. 2018. Disponível em: <<https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>>. Acesso em: 16 nov. 2019.

KELSEN, Hans. **Teoria Pura do Direito**. São Paulo: Martins Fontes, 1999. Tradução de João Baptista Machado.

KUHN, Thomas. **A função do dogma na investigação científica**. Curitiba: Ufpr, 2012. Tradução de Jorge Dias de Deus.

MAHAPATRA, Sambit. **Why Deep Learning over Traditional Machine Learning?** 2017. Disponível em:

<<https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>>. Acesso em: 01 nov. 2018.

MIKOLOV, Tomas et al. **Efficient Estimation of Word Representations in Vector Space**. Mountain View, Ca: Arxiv:1301.3781 [cs.cl], 2013.

NUNES, Dierle; LUD, Natanael; PEDRON, Flávio. **Desconfiando da imparcialidade dos sujeitos processuais**. Salvador: Juspodivm, 2018.

O'NEIL, Cathy. **Weapons of math destruction**. New York: Broadway Books, 2016.

O'NEIL, Cathy. **Weapons of Math Destruction**: Vídeo-conferência. Personal Democracy Forum 2015. Disponível em: <https://www.youtube.com/watch?v=gdCJYsKIX_Y>. Acesso em: 16 nov. 2019.

PENNINGTON, Jeffrey; SOCHER, Richard; MANNING, Christopher. Glove: Global Vectors for Word Representation. **Proceedings Of The 2014 Conference On Empirical Methods In Natural Language Processing (emnlp)**. Doha, Qatar, p. 1532-1543. out. 2014. DOI: 10.3115/v1/D14-1162 Disponível em: <<https://www.aclweb.org/anthology/D14-1162/>>. Acesso em: 16 nov. 2019.

PUGET, Jean François. **What Is Machine Learning?** 2016. Disponível em: <https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en>. Acesso em: 01 nov. 2018.

ROSA, Alexandre Morais da. **Guia do Processo Penal Conforme a Teoria dos Jogos**. 5. ed. Florianópolis: Emias, 2019.

ROSA, Natalie. **Ex-campeão de Go disse ter se aposentado por ser impossível vencer a IA**. 2019. Disponível em: <<https://canaltech.com.br/inteligencia-artificial/ex-campeao-de-go-disse-ter-se-aposentado-por-ser-impossivel-vencer-a-ia-156451/>>. Acesso em: 30 nov. 2019.

RUSSEL, Stuart; NORVING, Peter. **Artificial Intelligence: a modern approach**. 3. ed. Harlow (UK): Pearson Education Limited, 2014.

SAMUEL, Arthur. **Some Studies in Machine Learning Using the Game of Checkers**. IBM Journal of Research and Development, 2010.

SANTAELLA, Lúcia. **O que é semiótica**. São Paulo: Editora Brasiliense, 1983.

SIMON, Henrique Smidt. **Direito, Hermenêutica e Filosofia da Linguagem**. Belo Horizonte: Argumentvm, 2006. 148 p. (ISBN-10: 8598885096).

STANKOVIC, Mirjana; GUPTA, Ravi; ROSSERT, Bertrand Andre. **Exploring legal, ethical and policy implications of artificial intelligence**. 2017. Disponível em: <<http://www.globalforumijd.com/new/sites/default/files/documents/resources/Artificial-Intelligence-White-Paper-Draft-5Oct2017.pdf>>. Acesso em: 16 nov. 2019.

WARAT, Luis Alberto. **O Direito e sua linguagem**. 2. ed. Porto Alegre: Sergio Antonio Fabris (editor), 1995.

WITTGENSTEIN, Ludwig. **Tractatus Logico-Philosophicus**. São Paulo: Editora da Universidade de São Paulo, 1968. Tradução de José Arthur Giannotti. Disponível em: <<https://marcosfabionuva.files.wordpress.com/2011/08/tractatus-logico-philosophicus.pdf>>. Acesso em: 16 nov. 2019.

WITTGENSTEIN, Ludwig. **Investigações Filosóficas**. São Paulo: Nova Cultural, 1999. Tradução de José Carlos Bruni.

WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. **Vieses da justiça: como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva**. Florianópolis: Empório Modara, 2018.