

A TRANSPARÊNCIA ALGORÍTMICA

As ontologias advêm da Filosofia e se referem a um sistema de categorização dos objetos do mundo para a organização da realidade, o que, para o presente tema dessa tese, torna-se indispensável. Assim, as ontologias especificam conceitos sobre um dado campo de conhecimento e das relações, das estrições e dos axiomas válidos entre esses conceitos e suas instâncias. Isso prove uma boa taxonomia tanto de entidades quanto de instâncias desse universo de discurso.¹

Com isso, antes de adentrar-se no terreno fértil da aplicação Inteligência Artificial no Direito, por cautela, deve-se buscar um conhecimento mínimo da construção cognitiva tecnológica, esta que se conecta com o direito dia após dia. E por razões óbvias, não se pode desprezá-la, posto o risco de se ficar sem a compreensão dos fenômenos que inevitavelmente irão de surgir no Direito.

A Inteligência Artificial (IA),² que merece um capítulo dedicado nesta tese, é uma subárea da Ciência da Computação e, por isso, é de grande importância desvelar-se os conceitos sobre algoritmos, *machine learning*, redes neurais e *deep learning*, *malwares* e, indiscutivelmente, as diretrizes ética na construção dos sistemas de IA. Ao fim e ao cabo, são elementos estruturantes da Inteligência Artificial. Para que se tenha uma visão de mundo adequada para identificar-se e reconhecer-se a disruptura no direito, o primeiro capítulo, intitulado *Os elementos estruturantes da Inteligência Artificial*, está dividido em seis subcapítulos. São eles: (i) a transparência algorítmica; (ii) *machine learning* e os tipos de aprendizagem; (iii) as redes neurais artificiais e *deep learning*; (iv) a inteligência artificial e suas características; (v) a inteligência artificial contra *malwares*; e, por fim, (vi) as diretrizes éticas para construção coerente e integra dos sistemas de inteligência artificial. Quanto ao *review* conceitual, entende-se

¹ GUARINO, Nicola. Formal ontology and information systems. *In: Proceedings of FOIS*, v. 98, 1998, p. 81–97. Disponível em: https://www.researchgate.net/publication/272169039_Forma_l_Ontologies_and_Information_Systems. Acesso em: 09 nov. 2020.

² Na fase inicial do projeto Vitor, ainda não colocada em funcionamento no ano de 2019, o robô terá a função de aumentar a velocidade de tramitação dos processos por meio da utilização da tecnologia para auxiliar o trabalho do Supremo Tribunal. A Corte espera que, em breve, todos os tribunais do Brasil poderão fazer uso do Victor para pré-processar os recursos extraordinários logo após sua interposição, o que visa antecipar o juízo de admissibilidade quanto à vinculação a temas com repercussão geral, o primeiro obstáculo para que um recurso chegue à Suprema Corte.

necessário para compreensão terminológica conceitual da IA ao analisar-se o modo como a IA vem sendo inserida no Direito.

De certo modo, é razoável afirmar que a Inteligência Artificial, devido a sua incomparável característica transdisciplinar que navega nos diagnósticos médicos, com informações até então confessadas ao médico do paciente, ou até a outras aplicações autônomas de aeronaves, veículos, embarcações e espaçonaves,³ está em voga em diversas áreas do conhecimento. O direito está a recepcioná-lo em diversas tarefas repetitivas de reduzido grau intelectual.

No entanto, o avanço da Inteligência Artificial em áreas do conhecimento diferentes da Ciência da Computação exige a aproximação com o *expert* e o estudo da área do conhecimento que se objetiva cooperar. Do mesmo modo, para aproximar-se do Direito da Inteligência Artificial, coerentemente, deve-se desenvolver os conceitos fundamentais estruturantes da Inteligência Artificial. Indiscutivelmente, o mais relevante são as características oferecidas pelos algoritmos.

Importante frisar aos operadores do direito que o ensinamento de algoritmos é fundamental nos cursos da área tecnológica, por ser o passo inicial para o desenvolvimento do raciocínio lógico. E, por consequência, a introdução aos conceitos e prática da programação da Ciência da Computação⁴. O uso de algoritmos é muito mais intenso, complexo, profundo e impactante na rotina humana, indo desde a previsão do tempo até a seleção de currículos para uma vaga de emprego, uma organização de pautas de discussão social, bem como influências políticas. Os cientistas da biociência já auferiram que emoções não são meramente um fenômeno espiritual, mas concluíram que elas são algoritmos bioquímicos vitais para a sobrevivência e a reprodução dos mamíferos.⁵

³ A IA abrange uma enorme variedade de subcampos, desde áreas de uso geral, como aprendizado e percepção, até tarefas específicas como jogos de xadrez, demonstração de teoremas matemáticos e diagnósticos de doenças. RUSSEL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: a modern approach**. 2. ed. Nova Jersey, EUA: Prentice-Hall, 1998.

⁴ Para Wilson Engelmann o advogado, seu trabalho e o conteúdo com o qual lida, seja o conjunto normativo, sejam os fenômenos sociais que deverá tratar, estão em radical, talvez disruptiva forma de mudança. Perceber este “detalhe” e se preparar para trabalhar com ele será o grande desafio a ser enfrentado. ENGELMANN, Wilson; WERNER, Deivid Augusto. **Inteligência Artificial e Direito**. Ética, Regulação e Responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. (Coord.). São Paulo: Thomson Reuters Brasil, 2019. p.170

⁵ Um babuíno precisa de algoritmos muito mais complicados do que aqueles que controlam as máquinas de venda automática de opções de cafés e chás, pois necessita de cálculos corretos para um resultado extremamente importante: a própria sobrevivência. A probabilidade de o babuíno morrer de fome se não comer as bananas contra a probabilidade de o leão agarrá-lo.

Dito isso, um algoritmo pode ser definido como um conjunto de regras que define precisamente uma sequência de operações, para inúmeras e diferentes finalidades. A exemplo, pode-se citar os modelos de previsão de cenários e comportamentos;⁶ um conjunto de regras que cria o aprendizado de máquina por meio de um sistema computacional que busca realizar uma **tarefa T**, aprendendo a partir de uma **experiência E**, procurando melhorar uma **performance P**.⁷

Ainda, trata-se de um conjunto de instruções matemáticas ou regras que, se fornecidas especialmente a um computador, auxiliarão de modo eficiente e eficaz a calcular uma resposta com precisão para um problema.⁸ Contudo, no tocante desta pesquisa, demonstrar-se-á que há um problema incontornável nessa quadra da história dos sistemas de IA, isto é, esse conjunto de instruções matemáticas (*input*), apesar de construir o sistema de IA, não prestam *accountability* da decisão/resposta alcançada (*output*), na forma desejável ao Direito.

Os algoritmos corriqueiramente são utilizados para solucionar problemas do dia a dia. Eles são considerados comuns, mas também são criados

Para poder resolver esse problema, o babuíno deve levar em consideração alguns dados. A que distância estou das bananas? A que distância está o leão? Quão rápido sou capaz de correr? Quão rápido o leão é capaz de correr? O leão está acordado ou dormindo? Parece estar com fome ou saciado? Quantas bananas há ali? São grandes ou pequenas? Estão verdes ou maduras? Em acréscimo a esses dados externos, ele também deve considerar uma informação referente às condições de seu corpo. Se está faminto, faz sentido arriscar tudo pelas bananas, não importam quais sejam as probabilidades. Se, ao contrário, ele acabou de comer, e as bananas são só gulodice, por que assumir o risco? HARARI, Yuval Noah. **Homo Deus**: uma breve história do amanhã. Tradução de Paulo Geiger. 1. ed. São Paulo: Companhia das Letras, 2016. p. 93.

⁶ HANKE, Philip. **Algorithms and law**: a course on legal tech. Berna: University of Bern, LL.M. Institute of Public Law, 2018. Disponível em: <http://www.philiphanke.com/teaching.html>. Acesso em: 12 nov. 2019.

⁷ Basicamente, um algoritmo pode aprender a atingir um objetivo a partir de um grande volume de dados - suas **experiências**. Eis um exemplo para ficar mais claro. Suponha que nossa **tarefa** seja prever o resultado de um jogo de futebol. Como podemos fazer isso? Poderíamos fornecer ao computador dados sobre o técnico, composição do time, formação tática etc., seguidos dos resultados das partidas. Com um grande volume de dados em formato de pares (**variáveis, resultados**), esperamos que o computador possa aprender quais padrões levam à vitória. Também esperamos que o computador acerte mais conforme lhe apresentamos mais dados, pois ele terá mais exemplos de padrões que pode **generalizar** para situações ainda não vistas. Assim, quanto mais dados, isto é, quanto mais experiência, melhores serão os resultados, isto é, melhor será a nossa **performance**. HONDA, Hugo; FACURE, Matheus; YAOHAO, Peng. Os três tipos de aprendizado de máquina. *In*: LAMFO. Laboratório de Aprendizado de Máquina em Finanças e Organizações da UnB, 27 de julho de 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 15 nov. 2019.

⁸ ALGORITHM. *In*: CAMBRIDGE DICTIONARY. [Cambridge: Cambridge University Press], 2019. Disponível em: <https://dictionary.cambridge.org/dictionary/english/algorithm>. Acesso em: 13 nov. 2019.

especificamente para soluções precisas de *hards cases*, como de grande importância no desenvolvimento de soluções informáticas com as características específicas e determinadas. A exemplo, há: *input* dados fornecidos capazes de fornecer respostas objetivadas; *output* quantificável produzido; *definiteness* como espécie de instrução clara e inequívoca; *finiteness* encerrado após um número finito de etapas; e *effectiveness*, com cada instrução sendo básica o suficiente para ser viável e executada por uma pessoa.

Para exemplificar, suponha-se criar um algoritmo que busque ordenar um conjunto de números inteiros de um conjunto a partir de 1. O referido exemplo é utilizado por Horowitz para definir um algoritmo, para gerar o determinado comando a seguir: diante de um conjunto de números inteiros que se encontram classificados (57123486 – 17523486 – 12573486 – 12375486 – 12345786 – 12345786), busque encontrar o menor e insira-o na lista ordenada (12345678). Nota-se que o comando é uma pergunta; por isso, não é considerado ainda um algoritmo, mas caso alcance a ordem objetivada (12345678), demonstrando como os números foram armazenados e como ele construiu o resultado, resultará num algoritmo objetivado.⁹ Vale destacar aqui que o algoritmo não faz

⁹ **Two sorting algorithms** - Suppose we want to sort in increasing order a deck of n cards, numbered 1 through n . Here are two algorithms for doing this. In the **mergesort** algorithm, we start with n piles of one card each. We then take pairs of piles and merge them together, by repeatedly pulling the smaller of the two smallest cards off the top of the pile and putting it on the bottom of our output pile. After the first round of this, we have $n/2$ piles of two cards each. After another round, $n/4$ piles of four cards each, and so on until we get one pile with n cards after roughly $\log_2 n$ rounds of merging. Here's a picture of this algorithm in action on 8 cards: 5 7 1 2 3 4 8 6 - 57 12 34 68 - 1257 3468 – 12345678. Suppose that we want to estimate the cost of this algorithm without actually coding it up. We might observe that each time a card is merged into a new pile, we need to do some small, fixed number of operations to decide that it's the smaller card, and then do an additional small, fixed number of operations to physically move it to a new place. If we are really clever, we might notice that since the size of the pile a card is in doubles with each round, there can be at most $\log_2 n$ rounds until all cards are in the same pile. So the cost of getting a single card in the right place will be at most $c \log n$ where c counts the "small, fixed" number of operations that we keep mentioning, and the cost of getting every card in the right place will be at most $cn \log n$. In the "selection sort" algorithm, we look through all the cards to find the smallest one, swap it to the beginning of the list, then look through the remaining cards for the second smallest, swap it to the next position, and so on. Here's a picture of this algorithm in action on 8 cards: 57123486 – 17523486 – 12573486 – 12375486 – 12345786 – 12345786 – 12345687 – 12345678. This is a simpler algorithm to implement than mergesort, but it is usually slower on large inputs. We can formalize this by arguing that each time we scan k cards to find the smallest, it's going to take some small, fixed number of operations to test each card against the best one we found so far, and an additional small, fixed number of operations to swap the smallest card to the right place. To compute the total cost we have to add these costs for all cards, which will give us a total cost that looks something like $(c_1n + c_2) + (c_1(n - 1) + c_2) + (c_1(n - 2) + c_2) + \dots + (c_11 + c_2) = c_1n(n + 1)/2 + c_2n$. For large n , it looks like this is going to cost more than mergesort. But how can we make this claim cleanly, particularly if we don't know the exact values of c , c_1 , and c_2 ? ASPNES, James. **Notes on data structures and programming**

nenhum juízo de valor para além de sua programação; o *output* decorrerá exclusivamente dos *inputs*. Nesse processo, o algoritmo fornecerá uma resposta/decisão com base nos dados de entrada.

Com relação ao funcionamento algorítmico, pode-se classificá-los em duas espécies: os programados e os não programados. São os algoritmos programados que seguem as operações parametrizadas pelo programador. Assim, a informação ingressa no sistema, o algoritmo atua sobre ela e o resultado (*output*) sai do sistema. Nesse processo, o programador domina todas as etapas operacionais de construção algorítmica.

A ideia motriz dos algoritmos não programados ficou conhecida como aprendizado de máquinas denominada de *machine learning*, e exsurtiu, em 1950, no artigo *Computational Machinery and Intelligence*; porém, ganhou dinamismo somente após três décadas. Por meio das resoluções de equações e da análise de textos em diferentes linguagens, desenvolvia-se, na ocasião, uma capacidade randômica de aprendizado de máquinas.¹⁰

A aparição mais corriqueira dos algoritmos não programados, considerada a categoria primeira, é aquela que emprega algoritmos supervisionados,¹¹ quando o sistema é construído com dados elaborados e previamente escolhidos por seres humanos. Busca-se prever uma **variável dependente** a partir de uma lista de **variáveis independentes**. O aprendizado de máquina supervisionado é a área que concentra a maioria das aplicações bem-sucedidas e onde a maioria dos problemas já estão bem definidos.¹²

techniques. Yale: University of Yale, 2019. Disponível em: <http://www.cs.yale.edu/homes/aspnes/classes/223/notes.pdf>. Acesso em: 15 nov. 2019. p. 204.

¹⁰ TURING, Alan M. Computing Machinery and Intelligence. **Mind**, New Series, v. 59, n. 236, p. 433-460, out. 1950. Disponível em: <https://phil415.pbworks.com/f/TuringComputing.pdf>. Acesso em: 20 ago. 2019.

¹¹ Uma espécie de estruturação algorítmica que funciona de forma supervisionada são as redes neurais artificiais (com *back propagation*), a ser visitada no próximo subcapítulo 3.3, que inspiradas no cérebro humano, têm modelo de aprendizagem baseada em erros e acertos, com identificação paulatina dos caminhos e decisões mais corretas para atingir determinados objetivos.

¹² Ao longo do tempo, grandes provedores de tecnologia e até comunidades de código aberto disponibilizaram uma rica gama de bibliotecas e de ferramentas úteis para construção destes sistemas, de forma gratuita e aberta, por exemplo TensorFlow, PyTorch e scikit-learn. O TensorFlow é uma plataforma completa de código aberto para machine learning. Ele tem um ecossistema abrangente e flexível de ferramentas, bibliotecas e recursos da comunidade que permite aos pesquisadores levar adiante ML de última geração e aos desenvolvedores criar e implantar aplicativos com tecnologia de ML. TENSORFLOW. In: TENSORFLOW homepage. [S.l., 2020?]. Disponível em: <https://www.tensorflow.org/>. Acesso em: 26 out. 2020.

Quadro 2 – Relação de *inputs* x *outputs*

Variáveis Independentes	Variáveis Dependentes
Anos de Carreira, Formação, Idade	Salário
Idade do Carro, Idade do Motorista	Risco de Acidente Automotivo
Texto de um livro	Escola Literária
Temperatura	Receita de venda de sorvete
Imagem da Rodovia	Ângulo da direção de um carro autônomo
Histórico escolar	Nota no ENEM

Fonte: Elaborado pelo autor.

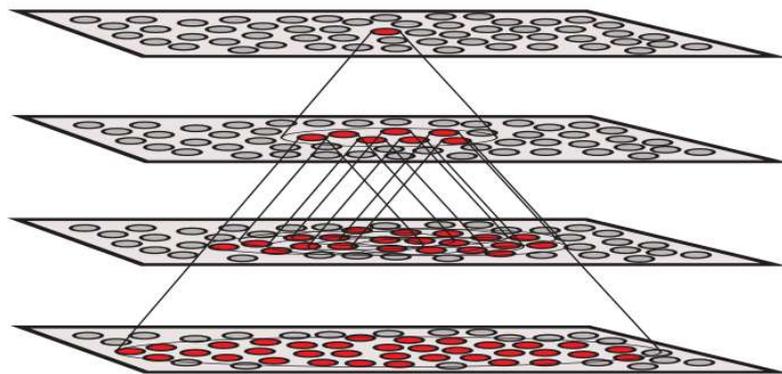
Observa-se que a característica básica de sistemas de aprendizado supervisionado é que os dados que são utilizados para treiná-los contêm a resposta desejada. Isto é, contêm a variável dependente resultante das variáveis independentes observadas, em que os dados são anotados com as respostas ou classes a serem previstas.

Nesse caso, a gama de dados rotulados e a saída objetivada são carregados e treinados no sistema, e o modelo ajusta as suas variáveis para mapear as entradas para a saída correspondente. O sistema é carregado com um objetivo mandamental *output*, bem como vários *inputs*, que serão testados em várias direções. Após a inserção dos *inputs*, constatar-se-ão inúmeros erros, que demandarão o retorno para nova programação (*error back propagation*). Já as tarefas mais coerentes passam a construir as camadas neurais internas (*hidden layers*). Isto é, na medida em que os algoritmos passam a conferir um

peso maior às conexões acertadas, os resultados desejados começam a dominar as tarefas.¹³ E nesse sentido, convém destacar tal especificidade que à frente será de grande valia para elucidações dos limites de aplicabilidade dos sistemas de IA no Direito, posto que tais correções de algoritmos enviesados no campo empírico, obstam os sistema de IA de aproximar-se das decisões judiciais. E, além disso, quanto maior o número de camadas de aprendizado, maior é a probabilidade de vieses, e diminui-se a capacidade do sistema de IA oferecer a *accountability* dos *outputs*.

Destaca-se que, dentre as técnicas mais conhecidas para resolver problemas de aprendizado supervisionado, estão: regressão linear, regressão logística, redes neurais artificiais, máquina de suporte vetorial (ou máquinas kernel), árvores de decisão, k-vizinhos mais próximos e Bayes ingênuo.¹⁴

Figura 1 – Camadas de aprendizado supervisionado



Fonte: Honda; Facure; Yaohao (2020).

Convém destacar aqui que os inúmeros erros (*vieses*) são resultantes da inserção dos *inputs*, o que demandam o retorno para nova programação (*error back propagation*). No caso do sistema de IA Victor, caso ocorra após a etapa

¹³ RUMERLHART, David E.; HILTON, Geoffrey E.; WILLINANS, Ronald J. Learning Representations by back-propagating errors. *Nature*, v. 323, n. 9, p. 533-536, out. 1986. Disponível em: https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf. Acesso em: 12 nov. 2019.

¹⁴ HONDA, Hugo; FACURE, Matheus; YAOHAO, Peng. Os três tipos de aprendizado de máquina. In: LAMFO. Laboratório de Aprendizado de Máquina em Finanças e Organizações da UnB, 27 de julho de 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 16 nov. 2019.

laboratorial, isso passa a ser constatado somente na prática, isto é, somente quando o sistema de IA já tiver realizado a seleção de uma tese de modo equivocado, o que se torna um problema por ora incontornável.

Os algoritmos não supervisionados (*non-supervised learning algorithms*) são considerados uma segunda categoria relevante, que são utilizados para descobrir padrões em determinado conjunto de dados não rotulados. Eis um objetivo em si mesmo ou um meio para atingir outro determinado objetivo. Em outras palavras, o algoritmo de aprendizagem busca encontrar estrutura nas entradas fornecidas de maneira independente. Sem qualquer classe pré-definida, os algoritmos não supervisionados têm a capacidade de organizar e de oferecer amostras desejáveis.

Exemplificando, são algoritmos utilizados pelas instituições bancárias para aprovar a concessão de empréstimos em que o desempenho dos algoritmos demonstra uma aparente precisão¹⁵. Assim, os dados serão referentes ao histórico do cliente, considerando o seu *score* de crédito. Em outras palavras, os algoritmos captam as características abonatórias e desabonatórias do tomador de crédito. Ainda, a referida técnica é empregada no reconhecimento e na identificação de faces e de vozes, além da criação de sistemas de tomada de decisão em curto espaço de tempo, viabilizando, por exemplo, a construção de carros e drones autônomos.

Ocorre que, em muitas ocasiões, o alcance de dados organizados é praticamente impossível, ou demanda um custo financeiro desproporcional ao resultado a ser alcançado.¹⁶ No caso da manipulação com os algoritmos não

¹⁵ Aparente precisão no sentido que, muitos modelos utilizam até mesmo dados das redes sociais do solicitante para o cálculo do *credit score*, baseados nas conexões sociais do indivíduo, o resultado vincula-se ao grupo social no qual o solicitante está inserido. Para exemplificar, refiro-me ao relatório do Federal Reserve apresentado no ano de 2007 ao Congresso dos Estados Unidos, que apontou os negros e hispânicos com um *credit score* inferior ao de brancos e asiáticos. “Differences in credit scores among racial or ethnic groups and age cohorts are particularly notable because they are larger than for other populations. For example, the mean normalized TransRisk Score for Asians is 54.8; for non-Hispanic whites, 54.0; for Hispanics, 38.2; and for blacks, 25.6” Disponível em:

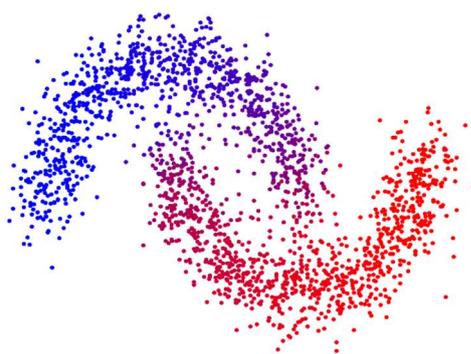
<https://www.transunion.com/resources/transunion/doc/compliance-and-legislative-updates/CreditScoreRpt>. Acesso em: 08 out. 2019.

¹⁶ Por exemplo, imagine que você é dono de um comércio e quer conhecer o perfil dos seus consumidores. Pode haver um perfil de consumidor que sempre compra vinho e queijo ou que compra carne e carvão ou ainda leite em pó e fralda. Se esse for o caso, colocar esses produtos em prateleiras distantes pode aumentar vendas, já que aumentará o tempo e o percurso do cliente no mercado. No entanto, nesse caso não estamos anotando para cada compra à qual o perfil do consumidor pertence. Mais ainda, sequer sabemos quantos perfis de consumidores há. HONDA, Hugo; FACURE, Matheus; YAOHAO, Peng. Os três tipos de aprendizado de máquina.

supervisionados, o computador terá de descobrir os perfis sem dados anotados e necessitará de métodos de aprendizado não supervisionados. A saída é uma espécie de predição, observando, nos registros de dados, se existem padrões que permitam a inferência de um grupo ou perfil a ser destacado. Outra opção seria ver diretamente quais dados são correlacionados e, então, aprender uma regra associativa entre eles.

Resulta afirmar que, com o aprendizado não supervisionado, pretende-se, ao fim e ao cabo, encontrar uma representação dos dados mais informativa dos que aqui são dispostos. Essa representação exsurge mais informativa e mais simples, destacando a informação nos pontos mais relevantes. A exemplo, os dados oriundos das palavras de um texto jurídico e a sua forma representativa seriam a representação matemática das palavras.

Figura 2 – Representação de dados do aprendizado não supervisionado



Fonte: Honda; Facure; Yaohao (2020).

Já no nosso cotidiano, os exemplos de aprendizado não supervisionados são sistemas de recomendação de filmes ou de músicas, detecção de anomalias e visualização de dados. A estruturação algorítmica que funciona de forma não supervisionada para atingir determinada finalidade é a rede neural convolucional, utilizada com sucesso no reconhecimento de imagens e no processamento de vídeo. Na área da saúde, a técnica é utilizada para o diagnóstico de determinadas doenças.¹⁷

In: LAMFO. Laboratório de Aprendizado de Máquina em Finanças e Organizações da UnB, 27 de julho de 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 16 nov. 2019

¹⁷ GARDNER. G. G. et al. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. **British Journal of Ophthalmology**, v. 80, n. 11, p. 940-944, nov.

As técnicas mais conhecidas para resolver problemas de aprendizado não supervisionado são: redes neurais artificiais, Expectativa-Maximização, clusterização k-médias, máquina de suporte vetorial (ou máquinas kernel), Clusterização Hierárquica, *word2vec*, análise de componentes principais, florestas isoladoras, mapas auto-organizados, máquinas de Boltzmann restritas, *eclat*, *apriori* e *t-SNE*.¹⁸

O modelo de aprendizado não supervisionado está na fronteira do conhecimento, pois são considerados mais complicados do que problemas de aprendizado supervisionado, precipuamente porque não se tem a resposta prévia nos dados. Como consequência disso, fica extremamente complicado e controverso avaliar um modelo de aprendizado não supervisionado.

E aqui está o cerne da questão do sistema de IA Victor: o problema da incapacidade de antever as tarefas desempenhadas pelos algoritmos de *machine learning*. Mesmo com projetos de sistemas de IA com acurácias elevadas, os aprendizados mais simples ou supervisionados não dispõem do racional matemático do seu funcionamento na sua totalidade.¹⁹ Em outras palavras, uma decisão sem *accountability*, o que dificulta muito sua atuação no Direito, em que qualquer decisão deve ser fundamentada.

No tocante, a transparência algorítmica pode ser entendida com base na acessibilidade e na compreensibilidade.²⁰ Com relação ao sistema de IA dar maior ênfase para a primeira, isto é, para a publicização do código-fonte, o ponto crucial se refere ao componente da compreensibilidade. Com a estrutura do aprendizado de máquina se tornando mais complexa, a mera abertura do código-

1996. Disponível em: www.ncbi.nlm.nih.gov/pmc/articles/PMC505667/pdf/brjophthal00011-0006.pdf. Acesso em: 16 nov. 2019.

¹⁸ HONDA, Hugo; FACURE, Matheus; YAOHAO, Peng. Os três tipos de aprendizado de máquina. *In*: LAMFO. Laboratório de Aprendizado de Máquina em Finanças e Organizações da UnB, 27 de julho de 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 15 nov. 2019.

¹⁹ FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik N. Arbitrium ex Machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. **Revista dos Tribunais Online**, v. 995, p. 1-16, set. 2018. Disponível em: <http://governance40.com/wp-content/uploads/2018/11/ARBITRIUM-EX-MACHINA-PANORAMA-RISCOS-E-A-NECESSIDADE.pdf>. Acesso em: 12 nov. 2019.

²⁰ MITTELSTADT, Brent Daniel et al. The ethics of algorithms: Mapping the debate. **Big Data & Society**, v. 3, n. 2, p. 1-21, jul./dez. 2016. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/2053951716679679>. Acesso em: 16 set. 2019.

fonte não auxilia na compreensão do *modus operandi* do sistema de IA,²¹ e aqui reside o problema incontornável ao Direito.

A opacidade nessa quadra da história é um problema que acompanha os especialistas da ciência da computação desde sempre. Em termos matemáticos, o funcionamento do aprendizado de máquina é invisível para todos; logicamente, salvo os matemáticos e os programadores. Desse modo, quando equivocados (enviesados), o veredito algorítmico se torna irrefutável, ou seja, imune de discordâncias e contestações. Isso resulta afirmar que o viesse poderá ser perpetuado e, assim, contribuirá, a exemplo, para as disseminações de eventuais discriminações e desigualdades sociais num efeito *feedback loop*.²²

Dessa forma, o código-fonte apenas demonstra a regra do *machine learning* utilizado e não a fundamentação decisória da seleção da questão de repercussão geral. Isso porque essa seleção nasce automaticamente a partir dos dados da rede neural artificial construída no sistema de IA, por isso, anteriormente se destacou que, quanto mais camadas das redes neurais profundas, menor é a capacidade de *accountability*.

Essa opacidade do aprendizado deriva exatamente da evolução da dimensionalidade do próprio sistema de IA²³. Ou seja, quanto mais julgados do STF transformados em algoritmos para construção do *machine learning*, mais opaca a explicação do racional da decisão dos sistemas de IA.²⁴ Destaca-se que essa opacidade que impede a identificação prévia dos algorítmicos enviesados,

²¹ A ideia de transparência algorítmica não se esgota na ideia de publicização da acessibilidade do código-fonte, pois deve demonstrar cabalmente a compreensibilidade do racional decisório, que faz referência ao efetivo entendimento de aspectos fundamentais de sua forma de operação (Ibid.).

²² O'NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. Nova York: Crown Publishers, 2016.

²³ Para Burrell, a opacidade dos sistemas de aprendizado é resultado da alta dimensionalidade de dados, da complexidade do código-fonte e da variabilidade da metódica da tomada de decisões. Uma vez que há a empregabilidade de milhares de regras, e suas predições estarem sujeitas probabilisticamente de modo complexo, pela velocidade no processamento das informações, e pela multiplicidade de variáveis operacionais, encontra-se muito além das capacidades humanas de apreender as estruturas decisórias que empreguem a técnica de *machine learning*. Por isso, a disponibilidade do acesso ao código-fonte é insuficiente para compreender o processo decisório do sistema de IA. MATTHIAS, Andreas. The responsibility gap: Ascribing responsibility for the actions of learning automata. **Ethics and Information Technology**, v. 6, n. 3, p. 175-183, nov. 2004. Disponível em: https://www.researchgate.net/publication/225249319_The_responsibility_gap_Ascribing_responsibility_for_the_actions_of_learning_automata. Acesso em: 17 set. 2019.

²⁴ MARTIJN, Van Otterlo. A machine learning view on profiling. In: HILDEBRANDT, Mireille; DE VRIES, Katja (Eds.). **Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology**. Abingdon: Routledge, 2013. p. 41-64.

que se diferenciam do viesse cognitivo, este que deriva da discricionariedade resultante do consequencialismo por razões políticas e econômicas. Portanto, os sistemas de IA no Direito, sobremaneira, devem preocupar-se com as duas espécies de enviesamentos algorítmicos, que são considerados como elementos determinísticos para limitar a aplicabilidade dos sistemas de IA no âmbito da seleção das questões de temas de repercussão geral.

Ainda, por fim, a terceira categoria de aprendizagem de máquinas é denominada de aprendizagem por reforço (*reinforced learning algorithms*). Nela, a máquina tenta aprender qual é a melhor ação a ser tomada. Dependendo das circunstâncias que essa ação será executada, os algoritmos de aprendizagem por reforço são treinados para tomar decisões. Há um *feedback* sobre o sucesso ou erro do *output*, que será utilizado para aprimorar o algoritmo.

Os algoritmos de aprendizagem por reforço, ao contrário dos algoritmos supervisionados e não supervisionados, não criam *outputs* corretos. No entanto, são os mais semelhantes ao comportamento dos seres humanos, pois focam na questão de maximizar a performance com base em consequências positivas ou negativas. É como um homem que se embriaga e, ao acordar no dia seguinte de ressaca, constata os efeitos negativos do álcool na corrente sanguínea. Os algoritmos de aprendizagem por reforço são de comum utilização em jogos, e a pontuação maior ou menor que eles atingem no processo funciona como recompensa.

A evolução humana sempre foi algo incerto e aleatório; dúvidas quanto à evolução da humanidade foi algo sempre perseguido pelo homem. E essa incerteza deve ser considerada e incorporada nas eventuais mudanças no ambiente do processo da melhor decisão. De fato, essa ideia de mudança no processo decisório caminha junto com o conceito de aprendizagem por esforço na área da psicologia, em que a decisão de um indivíduo a depender pode ser recompensada ou punida. Com repetições dos eventos, estima-se que o indivíduo passe a associar as ações que lhe garantam mais recompensa dentro de determinado ambiente que lhe é proposto. Esse objetivo na área da psicologia é denominado de *behaviorismo*.²⁵

²⁵ B. F. Skinner como um dos principais expoentes, um famoso psicólogo que, dentre outros experimentos, usou a ideia de recompensas e punições para treinar pombos para conduzir mísseis na Segunda Guerra Mundial.

Essa ideia de recompensa figura no aprendizado de máquinas, sendo que a máquina, diante de um conjunto de cenários possíveis, acaba escolhendo uma determinada ação a se tomar e recebendo a recompensa associada a essa ação ótima. Assim, quando maximizada a decisão oriunda do processo de aprendizado, a máquina recebe a recompensa associada a essa ação específica, assim como, primordialmente, também recebe e memoriza essa combinação para atos futuros. O processo se repete, passa por diversos erros (*error back propagation*), até que a máquina seja capaz de escolher a melhor ação e decidir, do mesmo modo, o futuro num caso idêntico. Diferentemente do ser humano, que, submetido a processos repetitivos, tende a aumentar a probabilidade de falhas, o algoritmo é o caminho que possibilita a máquina a gravar e a jamais esquecer a combinação ótima já alcançada em cenários pretéritos. Contudo, a problemática reside nos vieses que exsurtem somente na prática.²⁶

De modo que se possa aproximar a função e o desempenho algorítmico da realidade, utiliza-se o exemplo da causa-efeito no mundo animal. No adestramento animal, a punição e a recompensa são meios pelos quais se busca o sucesso do adestramento. Sabe-se que qualquer animal nasce sem obediência ao seu proprietário e, de início, os comandos desatendidos serão punidos com expressões negativas e, lamentavelmente, até com chicoteadas. Porém, quando o animal inicia o processo de aprendizado e demonstra seus primeiros atos de obediência, as recompensas são realizadas com afagos e alimentos de preferência do animal. Esse processo se repete inúmeras vezes até que o animal passe a sopesar a causa-efeito e a consubstanciar unicamente os atos que gerem recompensas.²⁷

²⁶ No ano de 2018, ocorreu um acidente com o carro autônomo da Uber, que atropelou a ciclista Elaine Herzberg, que pedalava à noite por uma estrada na cidade de Tempe, no Arizona. De acordo com as investigações, diversos fatores contribuíram para o acidente, mas o principal deles é que o motorista designado para o veículo — que precisa estar sempre atento na pista para assumir o controle caso perceba que a IA do carro cometerá um erro — estava assistindo a vídeos da internet ao invés de prestar atenção na estrada. Observem, que o erro algorítmico reconhecia um pedestre na estrada, mas não reconheceu na prática um pedestre na bicicleta.

²⁷ O famoso experimento do “cão de Pavlov” ilustra bem esse paradigma de aprendizagem. Ivan Pavlov foi um cientista russo notório por apresentar a ideia do “reflexo condicionado”, baseado no seguinte experimento: apresentando um pedaço de carne a um cão, o animal passa a salivar, desejando o alimento. Em vez de apresentar apenas a carne, Pavlov soava uma campainha sempre que isso acontecia; com a repetição, o cão passava a associar os dois “estímulos” (carne e campainha) e salivar assim que ouve a campainha.

A relação de causa-efeito é muito ilustrativa para se trazer para a ceara da ciência da computação, pois o algoritmo é muito eficiente para captar as combinações ótimas de causa-efeito. Exemplificando, no caso do mercado financeiro, há a possibilidade de se estruturar um portfólio no mercado financeiro e de ajustar a combinação de ativos comprados/vendidos a depender do retorno monetário (recompensa). Com isso, de alcançar ganhos de acordo com o mercado.

Outro exemplo ocorre com os aplicativos de trânsito: no momento de uma colisão, o algoritmo do sistema capta o congestionamento com a diminuição de velocidade para aquela parte do trajeto e deixa de receber as recompensas, acusando, assim, uma ocorrência atípica para os demais condutores. Então, há o deslocamento dos condutores para outra rota mais rápida, e isso se repete noutros incidentes no trânsito. O sistema algorítmico vai aprender que deve mudar os condutores para uma rota alternativa.

O desafio do sistema algorítmico, utilizando-se o exemplo do cão, é descobrir o lugar ótimo entre o comando *easy* “*exploration*” (fazer o cão sentar-se) e o comando *hard* “*exploitation*” (fazer o cão recuar, saltar, atacar etc.), que denota para aplicações extremamente complexas. A exemplo, eis o caso de identificar jurisprudências ou precedentes de um determinado tribunal para a aplicação num caso concreto.

Deve-se destacar que a parcela da crítica dos operadores do direito, quanto ao avanço da Ciência da Computação no Direito, reside no fato de que há possíveis escapes no controle sobre os processos de aprendizagem de algoritmos. Além disso, os avanços foram inspirados na biologia e na psicologia do cérebro humano.²⁸ Por isso, a doutrina de Desai sinaliza uma necessária regulação, para que não se desperdice a oportunidade de convergência dessas duas áreas.²⁹

O pesquisador Desai destaca um mal-entendido em torno dos algoritmos, que justifica a transparência como um conceito importante. Contudo, uma ilusão

²⁸ ITO, Joi; HOWE, Jeff. **Whiplash**: How to survive our faster future. New York/Boston: Grand Central, 2016. p. 240-241.

²⁹ DESAI, Deven R.; KROLL, Joshua A. Trust but Verify: a guide to Algorithms and the Law. **Harvard Journal of Law & Technology**, v. 31, n. 1, p. 1-64, abr. 2017. Disponível em: <https://jolt.law.harvard.edu/assets/articlePDFs/v31/31HarvJLTech1.pdf>. Acesso em: 17 nov. 2019

de clareza pode se configurar quando ocorre uma pura exposição algorítmica, com a simples disponibilidade de códigos-fonte.³⁰ Por isso, há uma preocupação importante que paira em torno da regulamentação sobre as aplicações algorítmicas em *softwares* e na evolução algorítmica.³¹ Também há uma preocupação em criar instrumentos que detectam desvios e utilizações errôneas, construindo uma espécie de sistema de auditoria e governança, exatamente na linha de uma necessária regulamentação.

Contudo, nesse ponto específico, já se desvelou que o (*error back propagation*), por mais que sejam disponibilizados os códigos-fonte ou criados os sistemas de auditoria e governança, os vieses serão verificados somente no funcionamento do sistema de IA. Além disso, os referidos códigos-fonte ou sistemas de auditoria e governança servem apenas para que os idealizadores do projeto possam revisar a performance do sistema de IA, uma espécie de *accountability* privada.

Convém destacar que, quando os algoritmos são relacionados com o Direito, torna-se indispensável, diante da característica da previsão e da especificação da finalidade objetivada, realizar a diferença entre os sistemas jurídicos baseados na *common law* e na *civil law*,³² de modo que se opte o sistema acertado para exploração algorítmica.

No estudo subsidiado e orientado por dados, a exemplo, torna-se tangível a possibilidade de prever certos resultados e, com o auxílio da *machine learning*, analisar redes, processar linguagem natural, mas, notadamente, afastando-se das decisões judiciais pelos limites de aplicabilidade dos sistema

³⁰ HARTMANN, Fabiano Peixoto; SILVA, Roberta Zumblick M. da. **Inteligência artificial e direito**. Curitiba: Alteridade Editora, 2019. p. 73.

³¹ O pesquisador Stuart Russel alerta para o fato que as máquinas evoluem tão rapidamente que criar algoritmos melhores é uma perda de tempo, e fundamenta com base na Lei de Moore. Qual é a lei de Moore? "Lei de Moore" refere-se a várias observações e previsões relacionadas a respeito do crescimento exponencial na densidade e/ou desempenho de circuitos eletrônicos. Um resumo moderno e útil, que não é fiel às declarações originais de Moore, é que o número de operações por segundo, por dólar gasto, dobra a cada N meses, onde N é aproximadamente 18. Alerta Stuart Russel quanto à equívocos comuns, pois a lei de Moore é uma lei da física. De fato, é uma observação empírica sobre o progresso da tecnologia; nada exige que ele continue e, é claro, não pode continuar indefinidamente. Já, os aumentos na velocidade atingiram um patamar, e as melhorias atuais em preço / desempenho vêm do aumento do número de núcleos (unidades de processamento) em um único chip. As máquinas estão ficando mais rápidas tão rapidamente que criar algoritmos melhores é uma perda de tempo. De fato, melhorias simples nos algoritmos costumam ser muito mais significativas do que melhorias no hardware. RUSSEL, Stuart. **Q&A: The Future of Artificial Intelligence**. In: University of Berkeley, 2016. Disponível em: <https://people.eecs.berkeley.edu/~russell/temp/q-and-a.html>. Acesso: 13 jun. 2020.

³² HARTMANN; SILVA, op. cit., p. 70.

de IA no Direito. Há possibilidades de traduzir um conjunto de regras, transformando-o em códigos executáveis pela máquina, naquilo denominado de *LegalTech*.³³

Atualmente, o fenômeno da aproximação do Direito com a tecnologia é um fenômeno mundial³⁴. Nota-se que muitas tarefas antes desempenhadas por humanos operadores do direito passaram a ser, de modo mais eficiente, desempenhadas por um *software*. O processo se tornou eletrônico; a Suprema Corte, que já analisava os possíveis temas de repercussão geral por meio do Plenário Virtual, com o advento da Pandemia do vírus SARS-CoV-2 causador de COVID-19, passou a fazer as sessões de plenário por intermédio da tecnologia de vídeo conferência, como se os ministros estivessem no plenário. Contudo, sempre se tem de ter em mente os efeitos colaterais da inserção tecnológica no Direito.

A fusão do Direito com a tecnologia ocorreu definitivamente durante a pandemia, quando muitos especialistas da área tecnológica passaram a afirmar que o avanço tecnológico de 8 a 10 anos se deu em 3 meses de pandemia. Assim, a corrida aos dados para construção de um aprendizado de máquina se tornou objeto de desejo de desenvolvedores de sistema de IA para o Direito.

Dito isso, nesse capítulo foi destacada a opacidade algorítmica e ausência de *accountability*, em especial, das redes neurais artificiais profundas que limita os sistemas no Direito. No próximo subcapítulo, estudar-se-á o aprendizado de máquina, que é uma atividade complexa e que exige cuidado por parte dos programadores. Atualmente, já se sabe que não é a quantidade de informações enviadas para a construção dos sistemas de IA que importa e, sim, a qualidade destas, pois dados enviesados – discricionários – também ensinarão a máquina a desempenhar suas funções de forma enviesada, ancorando, de forma automatizada, o consequencialismo por razões econômicas e políticas visitadas no capítulo anterior.

Nesse sentido, é importante visitar-se o modo de construção do aprendizado de máquinas e seus limites de aplicabilidade no Direito.

³³ HANKE, Philip. **Algorithms and law**: a course on legal tech. Berna: University of Bern, LL.M. Institute of Public Law, 2018. Disponível em: <http://www.philiphanke.com/teaching.html>. Acesso em: 12 nov. 2019.

³⁴ Ibid.

