


Consequências e perspectivas da aplicação de inteligência artificial a casos penais


Consequences and prospects of the application of Artificial Intelligence to criminal cases

Gustavo Mascarenhas Lacerda Pedrina¹

Universidade de São Paulo – São Paulo/SP

gumascarenhas@hotmail.com

 <http://lattes.cnpq.br/4122752912642517>

 <https://orcid.org/0000-0003-1259-8428>

RESUMO: Este artigo tem como finalidade fazer uma análise da aplicação atual e das perspectivas do que se conhece popularmente por inteligência artificial ao direito penal. São analisadas duas vertentes do aprendizado da máquina (*machine learning*): a computação estatística e a análise preditiva. Discute-se as consequências e alternativas para o uso de técnicas de inteligência artificial diante dos avanços nas ciências do cérebro. Alguns pontos relevantes a respeito do tema são destacados para promover a discussão quanto a quais devem ser os próximos passos na área e quais as possíveis aplicações com o estado-da-arte da tecnologia atual.

PALAVRAS-CHAVE: Inteligência Artificial; Computação estatística; Análise preditiva; Cérebro.

ABSTRACT: *This article aims to make an analysis of the current application and perspectives of what is popularly known by artificial intelligence to criminal law. Two aspects of machine learning are analyzed: statistical computation and predictive analysis. It discusses the consequences and alternatives for the use of artificial intelligence techniques taken the advances in the brain sciences, some relevant points on the topic are highlighted to*

¹ Doutorando, Mestre e Bacharel em Direito pela Universidade de São Paulo (USP). Research Fellow no Charles Houston Institute da Harvard Law School (EUA). Assessor de Ministro no Supremo Tribunal Federal.

promote discussions about future scenarios of this area and possible current applications.

KEY-WORDS: *Artificial Intelligence; Statistical computation; Predictive analysis; Brain.*

SUMÁRIO: Introdução; 1. Cognição e Inteligência artificial. 2. Aplicação atual: computação estatística. 3. Aplicações viáveis da tecnologia existente. 4. O futuro: análise preditiva. 5. Conclusão. Referências.

INTRODUÇÃO

O direito penal está em constante adaptação à sociedade, procurando aliar a ciência que produz com as correntes filosóficas de cada momento. A tendência atual, não só nessa área do conhecimento, mas no mundo como um todo, está no aumento do uso da tecnologia, notadamente o que se convencionou chamar de inteligência artificial (IA). As novas ferramentas podem, de fato, criar sistemas de justiça mais balanceados, com resultados mais justos em suas decisões, mas é preciso analisar os resultados dessas inovações técnicas sob o ponto de vista científico, com todas as cautelas e os testes que a ciência recomenda. Esse exame pode indicar melhores aplicações às novidades tecnológicas.

A inteligência artificial se apresentará, em breve, como a solução para certos problemas da humanidade, inclusive na confecção de cenários de estudo da mente humana. Embora seja improvável que esta última variável substitua totalmente o homem, é possível que ela ofereça cenários que simulem a tomada de decisão do ser humano e, ao fazê-lo, ajude-nos a entender as possíveis razões da ação.

Não se trata propriamente de desvendar o dilema do livre-arbítrio do ponto de vista médico, mas os avanços da tecnologia podem deixar no passado conceitos puramente eticizantes, baseados em crenças histórico-religiosas.

Para o processo penal, importa esclarecer que não há no seu uso a solução singular para o estabelecimento de procedimentos judiciais que levem à aferição da culpa de modo automático e equânime. Ainda

assim, é necessário esclarecer (1) se a máquina é capaz de substituir o jogador humano – ou se o será num futuro breve; (2) em que medida a tecnologia já existente pode ser utilizada e em quais seriam aplicações viáveis ao processo penal; (3) quais as perspectivas futuras de sua aplicação para a matéria.

É preciso tratar a tecnologia como uma técnica não-neutra e ainda em evolução, bem assim apresentar o real ponto de seu desenvolvimento, afastando mitos e versões não validadas de uso.

1. COGNIÇÃO E INTELIGÊNCIA ARTIFICIAL

A inteligência da máquina fascina há um bom tempo a humanidade. O conceito de Inteligência Artificial (AI na sigla em inglês) é atribuído ao cientista da computação JOHN MCCARTHY, que cunhou o termo em 1955. Pode-se resumir o conceito como a projeção de uma rede computacional para executar um conjunto definido de ações e aprender com a experiência.² ALAN TURING apresentou trabalhos seminais para a inteligência artificial ainda na década de 1950, que indicavam que a máquina poderia ser programada para aprender pelo mimetismo da inteligência humana.³ No decorrer do século XX, o imaginário popular passou a traçar a mente, ainda sob influência da perspectiva dualista, como um computador do corpo – e se havia um “computador” no ser humano, o desafio seria reproduzi-lo artificialmente (há nisso uma herança da lógica defendida por DESCARTES, segundo a qual nenhum sistema puramente físico poderia pensar, raciocinar ou ser consciente).

Atualmente, assumimos que a disseminação da tecnologia, que leva à sensação de já haver a máquina atingido a inteligência própria, trouxe o paradigma da inteligência artificial quase perfeita. Embora comuns os programas de reconhecimento de rosto, de identificação de

² Stanford University. AI. Disponível em <https://www-cs.stanford.edu/memorial/professor-john-mccarthy> Acesso em 18.07.2019

³ “We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision.” TURING, A. M., *Computing machinery and intelligence*. In: *Mind*, Oxford, Vol. LIX, Outubro de 1950. P. 460.

imagens, objetos e vozes, tradução simultânea e até de assessoramento financeiro, nem tudo é inteligência artificial. Há diversos mitos quanto a matéria. Pode-se ligar a televisão, o rádio ou acessar a internet, uma pessoa se depare com propaganda que reivindica que tal aplicação é inteligente, o que não é exatamente verdade. Para ser realmente inteligente a máquina deve aplicar noções de contexto, o que, apesar de possível, ainda é insólito e apresenta um nível de desenvolvimento muito aquém do que o propagandeado.

O que quase todas as aplicações existentes fazem é computação estatística – que de fato faz parte das teorias de aprendizado por máquinas (*machine learning*), mas não é exatamente *inteligente*: o computador apenas oferece de maneira autônoma a melhor resposta estatística para a questão colocada.

Em primeiro lugar, é importante notar que o ser humano – mais bem acabado exemplo de inteligência natural e objeto do desejo de mimetismo – não é uma criatura estatística. Trata-se de uma espécie com particularidades que a tornam única, sendo a principal delas a linguagem. BERWICK e CHOMSKY sustentam que a linguagem é um sistema de organização de pensamentos e que apenas a espécie humana é capaz de fundir pensamentos em forma de linguagem gráfica.⁴ Esse processo de fusão é estruturado, buscando sempre o caminho neural mais curto para uma resposta concatenada. O arranjo dessa rede complexa é a resposta do sistema cognitivo, fruto da evolução humana.

Em segundo, deve-se notar justamente a capacidade cognitiva do ser humano, que, com auxílio da linguagem gráfica, é capaz de construir cenários no passado, que fazem sentido no presente e ajudam a projetar o futuro. A linguagem nos possibilita essa construção de panoramas, levando-nos ao próximo ponto de destaque: a capacidade de contar e

⁴ Os autores chamam o processo de “*Merge*” que aqui traduzimos, por aproximação, como “fusão”, mas que é algo ainda mais complexo que propriamente uma fusão, tratando-se de característica inata ao ser humano, que o diferencia enquanto espécie: é uma operação que combina duas expressões para gerar uma nova expressão mais complexa, sem que se modifique ou descarte as duas expressões originais. Ver mais em BERWICK, Robert C., CHOMSKY, Noam, *Why only us: language and evolution*. Cambridge: MIT Press, 2017. P.102

entender histórias (“*storytelling*”)⁵. WINSTON, diante dessas premissas, propõe a *Strong Story Hypothesis*, segundo a qual “os mecanismos que permitem aos humanos falarem, entenderem e recombinarem histórias separam a inteligência humana da de outros animais.”⁶

A combinação da linguagem e dos contextos, a partir de cenários históricos, forma a cultura. A cultura, por sua vez, é elemento fundamental na caracterização das próprias emoções humanas: BARRETT realizou experimentos que demonstram que as emoções humanas não podem ser tidas como padrões, elas variam de sociedade para sociedade e até de um indivíduo para outro⁷.

Essas premissas são importantes para que se adquira uma visão básica quanto a matéria: entregar a resposta média não é fazer justiça. Um sistema de inteligência artificial que não possa contextualizar os aspectos subjetivos do sujeito em análise entregará necessariamente uma resposta injusta. No processo penal, conforme consigna FENOLL, isto pode ser ainda mais delicado:

“No processo penal, a automação é mais complicada e perigosa. Um delito não é algo tão comum como um conflito civil e uma ameaça a condições pessoais – tanto do agressor quanto da vítima – que devem ser consideradas cuidadosamente.”⁸

⁵ Antes Winston propõe a *Inner Language Hypothesis*, segundo a qual “Human intelligence is enabled by a symbolic inner language faculty whose mechanisms support both story understanding and the querying of perceptual systems.” WINSTON, Patrick. *The Strong Story Hypothesis and the Directed Perception Hypothesis*. AAAI Fall Symposium Series, 2011. Artigo digital. Disponível em <https://dspace.mit.edu/bitstream/handle/1721.1/67693/Submitted.pdf?sequence=1&isAllowed=y> Acesso em 09.07.2019. P. 6

⁶ “The Strong Story Hypothesis: The mechanisms that enable humans to tell, understand, and recombine stories separate human intelligence from that of other primates” WINSTON, Patrick. *The Strong Story Hypothesis and the Directed Perception Hypothesis*. digital. Disponível em <https://dspace.mit.edu/bitstream/handle/1721.1/67693/Submitted.pdf?sequence=1&isAllowed=y> Acesso em 09.07.2019. P. 2

⁷ Cf. BARRETT, Lisa Feldman. *How emotions are made: the secret life of the brain*. Boston: Houghton Mifflin Harcourt, 2017.

⁸ “*En el proceso penal, la automatización es más compleja y peligrosa. Al fin y al cabo, un delito no es algo tan frecuente como un conflicto civil y suele estar*

De fato, é preciso notar que não há nada propriamente estatístico na aquisição de elementos culturais ou na formação da cognição. O ser humano pode até ser afeito a padrões, mas a cognição deve ser diferenciada da racionalidade: nem todas as tomadas de decisão do cérebro humano são fruto da escolha racional, mas todas são certamente tomadas de acordo com princípios cognitivos⁹, formados a partir de aspectos únicos do ser humano, muito distantes de serem mimetizados razoavelmente pela máquina. Por essa razão, não nos parece razoável a aplicação de qualquer programa que não seja capaz de *entender* as reações humanas para aplicações que de qualquer maneira impliquem no cerceamento da liberdade, ao menos no estado-da-arte atual da tecnologia. Ainda assim, como sugerimos no item 3, pode-se utilizar a tecnologia existente em situações isoladas no processo penal.

2. APLICAÇÃO ATUAL: COMPUTAÇÃO ESTATÍSTICA

O aumento dos estudos e aplicações a respeito da inteligência artificial levanta questões éticas a respeito dos limites do uso da tecnologia. No direito, é notável o dado que mais de 30 estados norte-americanos utilizem atualmente programas de inteligência artificial para sugerir aos juízes sentenças e fianças. Todos estes programas utilizados hoje são baseados em computação estatística, mais precisamente em um esquema chamado de EBS (*evidence-based sentencing*).¹⁰ Ocorre que apesar do nome, essas *evidências* são na realidade dados objetivos relacionados ao sujeito, como gênero e endereço.

lleno de circunstancias personales – tanto del reo como de la víctima – que deben ser consideradas cuidadosamente, incluso com más cuidado que el actual.” Cf. FENOLL, Jordi Nieva, Inteligencia artificial y proceso penal. Madrid: Marcial Pons, 2018. p. 36.

⁹ Isto tomando em conta um indivíduo responsivo e pleno de suas capacidades.

¹⁰ Cf. STARR, S., *Evidence-based sentencing and the scientific racionalization of discrimination*. Stanford Law Review, Setembro de 2013. Disponível em: <https://ssrn.com/abstract=2318940> Acesso em 16.07.2019. p.2

Recentemente, o caso *Wisconsin vs. Loomis* levou a questão até a Suprema Corte estadunidense¹¹. A defesa de Eric Loomis questionou o uso de um programa de inteligência artificial (o COMPAS – *Correction Offender Management Profiling for Alternative Sanctions* – produzido pela Equivant) por parte do Juízo para determinar a sua condenação a seis anos em regime fechado¹². Loomis queria ter acesso aos critérios que levaram o robô algorítmico a recomendar sua pena – a Suprema Corte negou o recurso, assentando tratar-se de segredo industrial.

O COMPAS leva em consideração 137 itens, não necessariamente ligados a questões atinentes ao delito possivelmente praticado, como (i) se há antecedentes penais na família do réu ou (ii) se há ocorrências relacionadas a álcool e drogas envolvendo a pessoa em julgamento¹³.

No estado da Virgínia, a utilização de algoritmos para estabelecer condenações já acontece há mais de dez anos. CALISKAN-ISLAM, BYRON, e NARAYAAN já demonstraram o perigo no seu uso com tal fim¹⁴: algoritmos são necessariamente programados e essa programação pode conter um erro de viés ideológico¹⁵. Os pesquisadores demonstraram que sentenças produzidas por robôs algorítmicos, em relação a nomes geralmente atribuídos a pessoas de ascendência africana, são

¹¹ U.S. Supreme Court. *Wisconsin vs. Loomis*. Disponível em: <https://www.supremecourt.gov/Search.aspx?FileName=/docketfiles/16-6387.htm> Acesso em 19.04.2019

¹² Disponível em <https://harvardlawreview.org/2017/03/state-v-loomis/> Ver também <http://www.nytimes.com/2005/01/02/magazine/sentencing-by-the-numbers.html>, <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html> e <https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html> Acesso em 19.04.2019.

¹³ FENOLL, Jordi Nieva, *Inteligencia artificial y proceso penal*. Madrid: Marcial Pons, 2018. p. 68

¹⁴ Cf. CALISKAN-ISLAM, A. BRYSON, J.J, NARAYAAN, A. *Semantics derived automatically from language corpora necessarily contain human biases*. 2016. Princeton: Princeton University. Disponível em: <http://randomwalker.info/publications/language-bias.pdf>. Acesso em 19.04.2019.

¹⁵ “Bias should be the expected result whenever even an unbiased algorithm is used to derive regularities from any data; bias is the regularities discovered.” (Ibid., p. 01).

comumente mais duras do que aquelas que contêm nomes tradicionalmente europeus^{16,17}.

Ocorre que a aprendizagem de máquinas envolve um estágio de pré-processamento para melhorar a qualidade dos dados de entrada¹⁸. Esse pré-processamento, também chamado de rotulagem, pode ser automático ou humano. Há, portanto, uma primeira porta de entrada a resultados viesados: se o dado for automatizado, as evoluções sociais serão absorvidas pelo algoritmo com atraso; se a programação for humana, ela dependerá das opiniões de seu programador. Além disso, se nesse processamento forem incluídos dados como raça, gênero e endereço, o algoritmo indicará, punições maiores para determinados grupos sociais já marginalizados.

Mas o maior problema é outro: por tratar-se de segredo industrial, é impossível saber com precisão como os itens levados em consideração são combinados (em que ordem e proporção, por exemplo) para sugerir sentenças ou fianças. Isso viola ao menos dois princípios caros ao direito

¹⁶ “We have shown that machine learning can acquire prejudicial biases from training data that reflect historical injustice. (...) We show for the first time that if AI is to exploit via our language the vast knowledge that culture has compiled, it will inevitably inherit human-like prejudices. In other words, if AI learns enough about the properties of language to be able to understand and produce it, it also acquires cultural associations that can be offensive, objectionable, or harmful. These are much broader concerns than intentional discrimination, and possibly harder to address. (...) Our results show that European-American names have more positive valence than African-American names in a state-of-the-art word embedding. That means a sentence containing a European-American name will have a higher sentiment score than a sentence with that name replaced by an African-American name. In other words, the tool will display a racial bias in its output based on actor and character names. We picked this example because the argument follows directly from our experiments on names. But our results suggest that other imprints of human racial prejudice, not confined to names, will also be picked up by machine-learning models.” (Ibid., p. 10-11).

¹⁷ Cf. MONAHAN, J; SKEEM, J. Risk Assessment in Criminal Sentencing. *Virginia Public Law and Legal Theory Research Paper*, n. 53. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2662082. Acesso em 19.04.2019.

¹⁸ Cf. KOTSIANTIS, S. B.; KANELLOPOULOS, D; PINTELAS, P. E. Data pre-processing for supervised learning. *International Journal of Computer Science*. *Jornal digital*. vol. 1, no. 2, 2006, p. 111-117.

penal e processual penal: o devido processo legal, já que é impossível contraditar dados sem que se saiba nem mesmo quais são, e a proporcionalidade, porque não se pode garantir que sejam realizadas sempre as mesmas operações.

O que esse tipo de programa faz não difere em nada daqueles de análise de risco geralmente empregados em operações financeiras, por exemplo. De acordo com dados atuariais, fornece-se uma sugestão de sentença para o Juízo que, em última análise, aplica superficialmente (i) a prevenção geral e (ii) busca inibir a reincidência¹⁹. Ambos os resultados, no longo prazo, tornam-se verdadeira profecia aos sujeitos-objeto da análise, já que não é comum que mudem, por exemplo, de gênero (e se o risco de homens adultos for maior para a reincidência em um determinado delito, o programa recomendará maior tempo de pena, por partir do pressuposto que o indivíduo precisa de mais tempo de *ressocialização* – este também um conceito discutível, aliás).

A aplicação ao processo penal da atual de inteligência artificial, baseada em computação estatística, concentra-se na sugestão de sentenças e nessa medida, entendemos, apresenta problemas intransponíveis, ao menos por hora, por violar princípios basilares à matéria.

3. APLICAÇÕES VIÁVEIS DA TECNOLOGIA EXISTENTE

Ainda que se ressalve a aplicação para o mais notável ato do processo, a sentença, é possível que se admita a utilização do que se chama de inteligência artificial para atos processuais diversos. Bons exemplos são (i) a aplicação de teses em tribunais a partir da leitura

¹⁹ “The trend is called “evidence-based sentencing” (hereinafter EBS). “Evidence,” in this formulation, refers not to the evidence in the particular case, but to empirical research on factors predicting criminal recidivism. EBS seeks to help judges advance the crime-prevention objectives of punishment by equipping them with the tools of criminologists—recidivism risk prediction instruments grounded in regression models of past offenders’ outcomes. The instruments give considerable weight to criminal history, which is already central to modern sentencing schemes.” Cf. STARR, S., *Evidence-based sentencing and the scientific rationalization of discrimination*. Stanford Law Review, Setembro de 2013. Disponível em: <https://ssrn.com/abstract=2318940> Acesso em 16.07.2019. p.1

automática de processos, (ii) o estabelecimento de standards probatórios (iii) a busca por evidências nos autos e (iv) as formulações de perguntas pelas máquinas.

No primeiro caso, é importante perceber que a máquina, por tecnologia já existente, pode “ler” processos. Trata-se de OCR, do inglês *Optical Character Recognition* (Reconhecimento Ótico de Caracteres), um método aplicado de maneira a permitir que o conteúdo escrito em um documento no formato de imagem seja reconhecido e transformado em um arquivo de texto editável. Desde o final de 2018, o Supremo Tribunal Federal tem aplicado a técnica nos recursos extraordinários e agravos em recursos extraordinários que chegam a Corte. O robô-algorítmico apelada de Victor interpreta os recursos e os separa por tema. Nos 27 temas mais recorrentes – que representaram aproximadamente 60%²⁰ dessas classes processuais em 2018 – o Victor é capaz de processar a devolução ao Tribunal de origem ante a aplicação de uma tese já aprovada pelo Supremo ou para sobrestá-lo até que decidida.

No segundo, é possível que, pela análise de casos análogos se defina a aplicação de *standards* probatórios. Um exemplo pode ser a quantidade de entorpecente que caracteriza a materialidade do crime de tráfico. A análise em bloco de processos com a aplicação de AI pode indicar a quantidade média que costuma indicar a traficância, auxiliando na definição de parâmetros ainda abstratos na jurisprudência.

No terceiro ponto, é possível treinar a máquina para isolar as evidências nos autos de modo a fornecê-las a investigação e a tomada de depoimentos, auxiliando nos testemunhos e diminuindo a presença de falsas memórias nas narrativas.

No quarto, pode-se instruir a máquina para fornecer perguntas ao Juízo de modo a estabelecer parâmetros mais neutros para as inquirições.

Em todas essas aplicações é necessário, contudo, que sejam públicos os critérios de estabelecimento dos algoritmos e os caminhos combinatórios estabelecidos pela IA, de modo a afastar qualquer possível viés.

²⁰ Disponível em: <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=388443> Acesso em 16.09.2019

4. O FUTURO: ANÁLISE PREDITIVA

Para WINSTON, a análise histórica é parte fundamental da formação cognitiva do ser humano. As histórias, segundo essa linha de pensamento, formariam, ao longo do amadurecimento do cérebro humano, os padrões sobre cada contexto social – a cultura do indivíduo (o que é também conhecido como “storytelling”).

Na inteligência cognitiva, é preciso que o sujeito possa analisar cenários e tomar escolhas ainda que não racionais. WINSTON e seu grupo de pesquisa elaboraram uma aplicação capaz de, a partir de histórias comuns à cultura americana (de Shakespeare aos irmãos Grimm), aprender o que leva a determinadas ocorrências, como, por exemplo, o *regicídio*.²¹ Partindo desses critérios, a máquina desenvolveu panoramas futuros e pode até mesmo identificar conceitos como *revanche* em histórias que não mencionavam a palavra. Isto prova que a análise preditiva consegue, se bem executada, identificar a intenção do agente, mesmo que esta não esteja explicitada.²² A este processo de exame de cenários futuros de

²¹ “With our Start Parser-enabled translator, we readily express the needed if-then rules in English. Flexibility illustrating examples follow, exactly as provided to Genesis. • If X kills Y, then Y becomes dead. • If X harmed Y and Y is Z’s friend, then X harmed Z. • X wanted to become king because Y persuaded X to want to become king. • Henry may want to kill James because Henry is angry at James. • If James becomes dead, then James cannot become unhappy. As the examples show, rules can be expressed as if-then sentences or because sentences, with or without regular names, and possibly with the modifiers may or cannot. May marks rules that are used only if an explanation is sought and no other explanation is evident. Cannot marks rules that act as censors, shutting off inferences that would otherwise be made. In the example, we do not become unhappy when we are dead, even though killing involves harm and harm otherwise causes the harmed to become unhappy. Reflection-pattern descriptions are a bit more complicated. Here are two versions of revenge. • Revenge 1: X and Y are entities. X’s harming Y leads to Y’s harming X. • Revenge 2: X and Y are entities. X’s harming Y leads to Y’s wanting to harm X. Y’s wanting to harm X leads to Y’s harming X.” WINSTON, Patrick. *The Strong Story Hypothesis and the Directed Perception Hypothesis*. AAAI Fall Symposium Series, 2011. Artigo digital. Disponível em <https://dspace.mit.edu/bitstream/handle/1721.1/67693/Submitted.pdf?sequence=1&isAllowed=y> Acesso em 09.07.2019. P. 4

²² WINSTON, Patrick. *The Strong Story Hypothesis and the Directed Perception Hypothesis*. Disponível em <https://dspace.mit.edu/bitstream/handle/1721.1/67693/Submitted.pdf?sequence=1&isAllowed=y> Acesso em 09.07.2019. P. 8

acordo com critérios estabelecidos no passado e no presente damos o nome da análise preditiva.

A máquina já é capaz de realizar operações do tipo “*who-knows-what*”, respondendo justificadamente perguntas sobre cenários de crimes e sobre as intenções dos sujeitos ao tomarem a ação, bem assim diferenciar os motivos de um crime. Veja-se o exemplo²³:

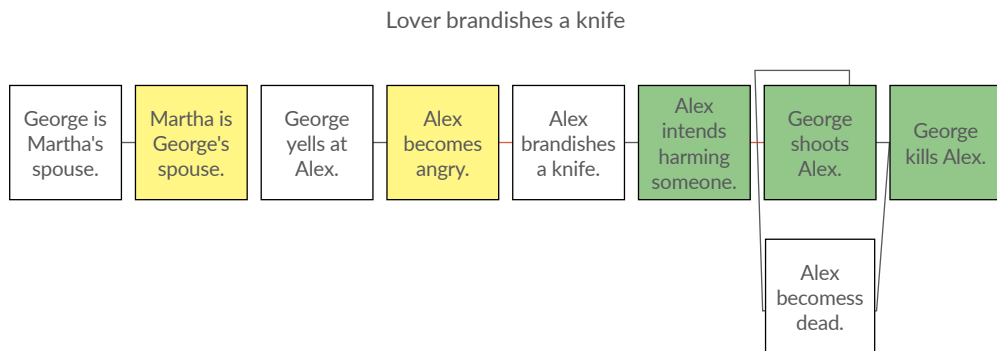


Figura 20: In the knife-brandishing version of a story raising a legal question, the elaboration graph indicates knife the brandishing is connected to the killing, suggesting *Self defense*.

²³ “Genesis can also compare two characters’ perspectives, attributing differences in interpretation to differences in what is observed. The following records an exchange between a human questioner and the Genesis: Why does Jean Valjean disagree with Inspector Javert? Inspector Javert and Jean Valjean disagree about “Jean Valjean is criminal”. Why does Jean Valjean think that Jean Valjean isn’t criminal? Jean Valjean infers that Jean Valjean isn’t criminal because [he] repents. Why did Inspector Javert think that Jean Valjean is criminal? Inspector Javert infers that Jean Valjean is criminal because [he] commits a crime. We believe that Genesis’s who-knows-what ability sheds light on our human ability to reason about what others know and believe. Genesis’s who-knows-what ability captures aspects of common sense (being within earshot, being unconscious or distracted, speaking over the phone or in another language), provides tools to aid in diplomacy and education (pinpointing differences in knowledge and experience), and suggests computational explanations of various psychological disorders (defects in mechanisms that enable understanding what others think).” WINSTON, Patrick, HOLMES, Dylan, *The Genesis enterprise: taking artificial intelligence to another level via computational account of human story understanding* Cambridge: MIT Computer Science and Artificial Intelligence Laboratory Center for Brains, Minds, and Machines, 2018. p. 28.

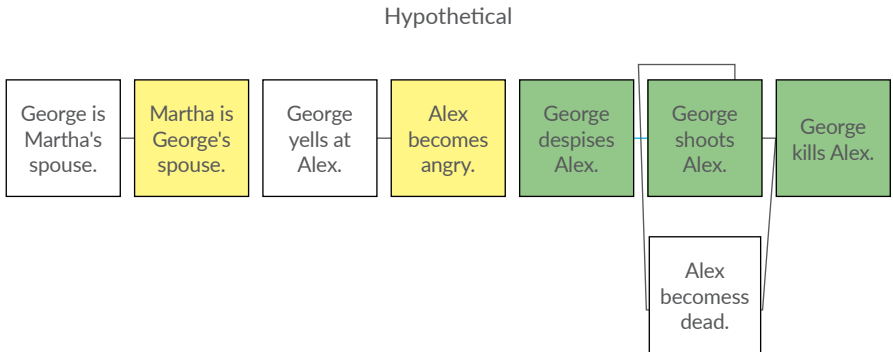


Figura 21: In the hypothetical version, with the brandishing removed, Genesis presumes the explanation for killing has to do with despising, suggesting guilt. At the concept level, *Self defense* becomes *Spiteful vengeance*.

RABINOWITZ ET AL. Alcançaram resultados semelhantes com a aplicação ToMnet²⁴, também baseada em técnicas de análise preditiva. Vemos nisso um caminho aplicável da inteligência artificial ao direito penal e processual penal no futuro: ao invés de tê-la como um instrumento de medição atuarial de risco dos indivíduos, tem-se a possibilidade de com ela se aferir as intenções e qual era a real capacidade cognitiva de uma pessoa no momento de tomada de decisão ao cometer um crime.

Por se tratar de sistemas baseados em *storytelling*, os programas de inteligência lastreados na análise preditiva tomam em conta os aspectos subjetivos do sujeito, interpretando-os de acordo com o cenário apresentado. Este tipo de uso da tecnologia oferece cenários preditivos com elevado grau de certeza que dão conta dos *porquês* de um sujeito haver

²⁴ “We therefore extended the ToMnet to be able to make declarative statements about agents’ beliefs. We achieved this by constructing a supervised dataset of belief states in the sample gridworld. We trained the UNREAL agents to report their beliefs about the locations of the four objects and the subgoal at every time step, alongside their policy. To do this, we added a head to the LSTM that output a posterior over each object’s current location on the grid (or whether it was absent). During training, the agents learned to report their best estimate of each object’s current location, based on its observations so far during the episode.” RABOWITZ, N. C., PEBERT, F., SONG, H.F., ZHANG, C., ESLAMI, S.M.A., BOTVINIK, M., *Machine Theory of Mind*. Disponível em <https://arxiv.org/abs/1802.07740> Acesso em 16.07.2019.

tomado certa decisão, podendo afetar, num futuro distante, até mesmo a culpabilidade – na medida em que pode entregar apreciação mais complexa da ação, vinculando a pessoa, de maneira individualizada, à conduta.

Esse tipo de utilização da tecnologia pressupõe, contudo, o aprendizado da máquina dos contextos em que cometido um delito e dos limites culturais envolvidos. Veja-se que não há ainda uma AI nesse nível de desenvolvimento, e que mesmo quando houver não será capaz de substituir o ser humano, cremos, na elaboração de sentenças ou atos que impliquem o cerceamento da liberdade do indivíduo. Poderá, ainda assim, fornecer ao julgador um cenário mais preciso do cometimento do crime, de modo a auxiliá-lo na tomada de decisão.

Pode-se reconhecer nesse tipo de aplicação uma melhor técnica para a instrução do processo-crime, a partir de parâmetros melhor delimitados, tornando mais rápido e preciso o processo penal.

CONCLUSÃO

Por se tratar de algo relativamente novo, ainda estamos aprendendo os reflexos das aplicações de inteligência artificial. Pode-se concluir do que estudado aqui que:

1. No ponto atual de desenvolvimento, a inteligência artificial não é comparável ao nível de desenvolvimento das redes neurais humanas, nem é capaz de substituir um julgador humano na determinação de atos que impliquem cerceamento da liberdade do indivíduo. A AI é desprovida de características básicas que fazem do julgador um ser humano, como emoções, capacidade de contextualização, cultura e linguagem.
2. Para se aliar, de forma válida, a inteligência artificial ao direito penal e processual penal os sujeitos não devem ser tratados de acordo com dados atuariais-objetivos. O uso de dados desse tipo em programas de inteligência artificial nos moldes em que ocorrem hoje nos EUA, para a sugestão de sentenças e fianças, podem aumentar a desigualdade social, levando ao encarceramento ainda mais concentrado de grupos já socialmente marginalizados.

3. É indispensável o estabelecimento de algoritmos claros e em acordo com preceitos constitucionais e de direitos humanos.
4. Há usos legítimos e ainda não explorados no processo penal para a AI no estado-da-arte atual, utilizando-a, por exemplo na (i) aplicação de teses em tribunais a partir da leitura automática de processos, (ii) estabelecimento de standards probatórios (iii) busca por evidências nos autos e (iv) formulações de perguntas pelas máquinas.
5. A técnica deve ser empregada sob o ponto de vista de garantia do indivíduo frente ao estado-juiz, nunca como elemento de prova pré-constituída de sua culpa. Trata-se, portanto, de meio de busca pela maior neutralidade possível no processo.
6. A análise preditiva é um caminho válido para o futuro da aplicação da inteligência artificial ao direito, estabelecendo cenários delituosos a partir de pontos de vista neutros que indiquem ao juízo um melhor nível de individualização da conduta. Para o processo penal especificamente, a utilização desse tipo de tecnologia pode resultar em maior precisão na instrução do processo-crime e melhores diretrizes decisórias ao julgador.

REFERÊNCIAS

BARRET, Lisa Feldman. *How emotions are made: the secret life of the brain*. Boston: Houghton Mifflin Harcourt, 2017.

BERWICK, Robert C., CHOMSKY, Noam, *Why only us: language and evolution*. Cambridge: MIT Press, 2017.

CALISKAN-ISLAM, A. BRYSON, JJ, NARAYAN, A. *Semantics derived automatically from language corpora necessarily contain human biases*. 2016. Princeton: Princeton University. Disponível em: <http://randomwalker.info/publications/language-bias.pdf>. Acesso em 19.04.2019.

FENOLL, Jordi Nieva, *Inteligencia artificial y proceso penal*. Madrid: Marcial Pons, 2018.

KOTSIANTIS, S. B.; KANELLOPOULOS, D; PINTELAS, P. E. Data preprocessing for supervised learning. *International Journal of Computer Science*. Jornal digital. vol. 1, no. 2, 2006, p. 111-117.

MONAHAN, J; SKEEM, J. Risk Assessment in Criminal Sentencing. *Virginia Public Law and Legal Theory Research Paper*, n. 53. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2662082. Acesso em 19.07.2019.

RABOWITZ, N. C., PEBERT, F., SONG, H.F., ZHANG, C., ESLAMI, S.M.A., BO-TVINIK, M., *Machine Theory of Mind*. Artigo digital: Cornell University, 2018. Disponível em <https://arxiv.org/abs/1802.07740> Acesso em 16.07.2019.

STARR, S., *Evidence-based sentencing and the scientific racionalization of discrimination*. Stanford Law Review, Setembro de 2013. Disponível em: <https://ssrn.com/abstract=2318940> Acesso em 16.07.2019

TURING, A. M., *Computing machinery and intelligence*. In: *Mind*, Vol. LIX, Outubro de 1950.

WINSTON, Patrick. *The Strong Story Hypothesis and the Directed Perception Hypothesis*. AAAI Fall Symposium Series, 2011. Artigo digital. Disponível em <https://dspace.mit.edu/bitstream/handle/1721.1/67693/Submitted.pdf?sequence=1&isAllowed=y> Acesso em 09.07.2019.

WINSTON, Patrick, HOLMES, Dylan, *The Genesis enterprise: taking artificial intelligence to another level via computational account of human story understanding*. Cambridge: MIT Computer Science and Artificial Intelligence Laboratory Center for Brains, Minds, and Machines, 2018.

YUDKOWSKY, E. The Ethics of Artificial Intelligence. In: FRANKISH, Keith; RAMSEY, William. *The Cambridge Handbook of Artificial Intelligence*. New York: Cambridge University Press, 2014.

Informações adicionais e declarações dos autores (integridade científica)

Declaração de conflito de interesses (conflict of interest declaration): o autor confirma que não há conflitos de interesse na realização das pesquisas expostas e na redação deste artigo.

Declaração de autoria e especificação das contribuições (declaration of authorship): todas e somente as pessoas que atendem os requisitos de autoria deste artigo estão listadas como autores; todos os coautores se responsabilizam integralmente por este trabalho em sua totalidade.

Declaração de ineditismo e originalidade (declaration of originality): o autor assegura que o texto aqui publicado não foi divulgado anteriormente em outro meio e que futura republicação somente se realizará com a indicação expressa da referência desta publicação original; também atesta que não há plágio de terceiros ou autoplágio.

Dados do processo editorial

(<http://www.ibraspp.com.br/revista/index.php/RBDPP/about/editorialPolicies>)

- Recebido em: 22/07/2019
- Controle preliminar e verificação de plágio: 23/07/2019
- Avaliação 1: 30/07/2019
- Avaliação 2: 05/08/2019
- Decisão editorial preliminar: 23/08/2019
- Retorno rodada de correções: 18/09/2019
- Decisão editorial final: 24/09/2019

Equipe editorial envolvida

- Editor-chefe: 1 (VGV)
- Editora-associada: 1 (CC)
- Revisores: 2

COMO CITAR ESTE ARTIGO:

PEDRINA, Gustavo M. L. Consequências e perspectivas da aplicação de inteligência artificial a casos penais. *Revista Brasileira de Direito Processual Penal*, Porto Alegre, vol. 5, n. 3, p. 1589-1606, set./dez. 2019. <https://doi.org/10.22197/rbdpp.v5i3.265>



Esta obra está licenciada com uma Licença *Creative Commons Atribuição-NãoComercial 4.0 Internacional*.