

## Fake news, inteligência artificial e processo penal

Os humanos não têm linguagem suficiente para demonstrar a Verdade (com "v" maiúsculo), na sua totalidade. Ela, porém, existe (para os que acreditam), mas escapa da ciência e, por evidente, aninha-se no campo da fé: da crença. Se é assim, da Verdade opera-se tão só com uma parte; e não com o todo. Assim, não é a Verdade. Disso, tem-se poucas dúvidas.



**Jacinto de Miranda  
Coutinho**  
Advogado e professor

O problema diz respeito à linguagem (que falta), mas que não retira a

pretensão de que se diga sobre a parte (uma verdade com "v" minúsculo), imaginando-se ou mesmo se querendo dizer sobre a Verdade. Diz-se uma mentira, assim. Para contornar o problema — e não raro seguir dizendo o que se não deve — faz-se uma diferença entre a Verdade (na qual é hoje pacífico que se não chega) e a verdade, à qual aparentemente se tem acesso pela parcialidade possível, entificada, e que usa como significantes — só para ficar no campo processual penal — o que se chama de "verdade formal", "verdade adjetiva", "verdade endoprocessual", e assim por diante. Diz-se sobre o ente, em geral pretendendo-se dizer sobre a Verdade; fala-se da parte se querendo referir ao todo; e isso não é possível, razão por que se mente. Nesse lugar, como se sabe, está a chamada verdade por correspondência, a forma mais lapidada de discursivamente se dizer uma pela outra. Perceba-se, então, que é um problema de linguagem; e não da coisa em si.

Apesar disso, segue-se falando a respeito da verdade, o que é ideológico e, por óbvio, uma questão de poder. Aqui se aninha a regra das regras ("*que abre o mais célebre código de todos os tempos* [\[1\]](#): *manter sob crença os sujeitos*"), expressa desde Justiniano e seu tratado de *Summa Trinitate*.



Assim, não se tem acesso à Verdade, mas, na parcialidade (garantida pela linguagem) é possível ter um consenso e, enquanto esse se sustenta, a verdade ganha uma estabilidade tanto discursiva quanto precária, pela via da aceitação social, a qual vai retroalimentando a racionalidade. Veja-se, como exemplo, a posição (atualíssima) de Glauco Giostra, ordinário de DPP na "*La Sapienza*", em Roma, falando da decisão no processo penal: "... *Impossível, porque não somos capazes de conhecer a verdade (...) ao final do qual um sujeito 'alheio' [terzo] chega a uma conclusão que a comunidade está disposta a aceitar como verdadeira...*" [2].

Uma mera verdade — que não é a Verdade — diz com uma mera mentira, como se fossem cara e coroa de uma moeda. O dilema (e aí o enorme problema) é que esse lugar oscilante do consenso, em ambas, abre espaço para a mentira desbragada, propositada, aquela das *Fake News* [3]. A falta da verdade (consensual) possibilita que a mentira seja assimilada como se verdade fosse, o que é terrível porque subverte o padrão, mormente no campo político, embora se espraie para os outros campos, inclusive o processo penal. Como qualquer jejuno percebe, elas colocam em risco a ordem constituída, as instituições e a própria democracia.

Sua produção é criminosa e elas provêm de pessoas que estão ali para tanto. Ao gerar uma ruptura na produção do consenso, há solo fértil para que as mentiras sejam assimiladas como se Verdadeiras fossem, permitindo que, em rebanho, cooptem-se os ingênuos ou mesmo aqueles predispostos a nelas acreditar, os quais passam a disseminá-las.

A necessidade de controle da disseminação das *Fake News* faz-se cada vez mais premente, mesmo porque, por um lado, com os instrumentos jurídicos conhecidos, é quase impossível realizá-lo. Basta ver que, pela ausência de controle no uso da *deep web*, o sucesso dos mal-intencionados é garantido. O volume absurdo de informações, juntamente com a construção de narrativas com pontos de verossimilhança, dificulta a percepção da mentira.

Por outro lado, pela capacidade de processamento rápido de dados, o uso da inteligência artificial (IA) vem sendo cotado para servir de mecanismo capaz de coibir, ou ao menos reduzir drasticamente, a propagação da desinformação. Seria uma solução interna, do próprio sistema.

Os programas como o ChatGPT e o Bard têm fascinado diversos usuários pela possibilidade de otimização na gestão de recursos e na disseminação de conhecimento. Ao mesmo tempo, têm suscitado interessantes e importantes questionamentos, tais quais os limites éticos do uso da IA, o estabelecimento de uma relação de dependência dos seres humanos à ferramenta, assim como a capacidade de substituição das pessoas pelas máquinas para a realização de tarefas [4].

Um ponto crucial para tratar do tema diz respeito às fontes que alimentam o sistema de IA, para a verificação da veracidade e da confiabilidade das informações. No caso do ChatGPT, para responder as perguntas dos usuários, o *software* recorre a um extenso banco de textos disponíveis na internet, que incluem de artigos científicos a frases colocadas em redes sociais, e utiliza da técnica de "Aprendizagem por Reforço com Feedback Humano", isto é, consegue extrair da interação com o indivíduo um *feedback*, para fornecer respostas de forma supostamente mais rápida, precisa e inteligente [5].



As críticas que vêm sofrendo dizem respeito à possibilidade de dar respostas plausíveis, mas incorretas, sem informar a incerteza ou o desconhecimento. A empresa NewsGuard, que rastreia a disseminação da desinformação online, realizou uma pesquisa perguntando à ferramenta questões relacionadas às *Fake News*, obtendo respostas contendo desinformação em 80% dos casos [6]. Ela poderia, então, ser instrumentalizada para a disseminação de *Fake News*, produzindo muito rapidamente variações convincentes e bem escritas de conteúdo de massa, sem informar as fontes. Como o *software* está em constante aprimoramento, ao tentar reproduzir o mesmo resultado da pesquisa, o New York Times obteve respostas com informações falsas em apenas 33% dos casos. Sobre o tema, informou a empresa responsável pelo ChatGPT que utiliza de máquinas e pessoas para monitorar o conteúdo que é inserido e produzido por ele, para identificar e filtrar dados falsos inseridos na plataforma [7].

O Bard, *software* em desenvolvimento pelo Google, utiliza uma linguagem de IA semelhante ao do ChatGPT, mas tem acesso à internet em tempo real e, combinando o *feedback* humano com testes internos, pretende "*garantir que as respostas (...) atendam a um alto nível de qualidade, segurança e fundamentação em informações do mundo real*", de forma responsável, evitando "*fazer declarações falsas que não sejam suportadas por fontes de informações externas*" [8]. A demonstrar preocupação com a instrumentalização da ferramenta para fins ilícitos e/ou antiéticos, a empresa, em 2018, publicou princípios para reger o desenvolvimento da IA

Como a linguagem e o consenso sobre o que é verdade estão em constante transformação, além da dificuldade de se classificar informações no sistema binário verdadeiro/falso, tal tarefa não parece ser simples e demanda um aprimoramento constante, com a colaboração entre as instituições para compartilhar conhecimentos e recursos para avançar no seu desenvolvimento responsável [9]. Afinal, diante dessas constantes mudanças, é crível que se encontrem meios de burlar a IA.

Por trás das empresas — é bom não esquecer — estão os homens e seus interesses econômicos [10], logo, *do dizer ao ser, no meio tem o infinito*. Porém, se, de fato, for possível controlar, ainda que parcialmente, a produção e disseminação das *fake news*, quem sabe se consiga restabelecer o procedimento de formação do consenso.

Os reflexos disso ao processo penal são diretos: permite-se restabelecer um mínimo de consenso entre os sujeitos do processo, de modo a impedir o julgamento por parâmetros criados em meio ao caos das *fake news* e a assegurar minimamente a equidistância entre as partes. Por outro lado, pode-se reduzir o excesso de subjetividade quando dos julgamentos, permitindo-se uma análise mais objetiva e científica da realidade quando da busca e avaliação da prova.

Não se ignora a complexidade da relação que se vai ter com a IA no mundo jurídico, sendo certo que a automatização de determinadas análises não pode ficar sujeita exclusivamente ao resultado dado pela ferramenta, especialmente porque podem reproduzir tendências, vieses e preconceitos, a perpetuar o sistema desequilibrado.

Porém, além de ser inevitável a integração da IA à realidade jurídica, uma análise percuciente de informações pode trazer à luz padrões de comportamento, bem como impedir, se bem manejada, a impregnação da desinformação no cotidiano, o que pode ser a salvação da democracia.



---

Há esperança, pelo menos até que se invente uma nova forma de burlar o padrão.

[1] LEGENDRE, Pierre. *O amor do censor: ensaio sobre a ordem dogmática*. Trad. Aluísio Pereira de Menezes *et al.* Rio de Janeiro: Forense Universitária-Colégio Freudiano, 1983, p. 19.

[2] GIOSTRA, Glauco. *Primeira lição sobre a justiça penal*. Trad. de Bruno Cunha Souza. São Paulo: Tirant lo Blanch, 2021, p. 14.

[3] Em interessante leitura da realidade, o jornalista Sean Illing identificou na campanha presidencial de Donald Trump, nos Estados Unidos, capitaneada por Steve Bannon, uma técnica que utiliza das redes sociais para manipular esse modo de formação do discurso, como estratégia de captação ou manutenção de poder político, não pela negação da verdade, mas sim por dificultar esse processo de encontrar a verdade. Inunda-se o ambiente com tanta desinformação que causa "um cansaço crescente sobre o processo de encontrar a verdade. E esse cansaço leva cada vez mais pessoas a abandonar a ideia de que a verdade é conhecível". O risco desse mecanismo, segundo o jornalista, é: "Se se interromper esse processo saturando o ecossistema com informações erradas e sobrecarregar a capacidade de mediação da mídia, poderá interromper o processo democrático", gerando o caos e interferindo na forma de criação do consenso. (ILLING, Sean. "Flood the zone with shit": How misinformation overwhelmed our democracy. *Vox Media*, 06/02/2020. Disponível em: <<https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation>>. Acesso em: 04/02/2023. Tradução de Luiz Campos Jr. disponível em: <<https://dagobah.com.br/inunda-a-area-com-merda-como-a-desinformacao-dominou-nossa-democracia/>>).

[4] Sobre o tema: STRECK, Lenio Luiz. Inteligência artificial: e o menino nunca mais foi visto no aeroporto. *Conjur*, 02.02.2023. Disponível em: <<https://www.conjur.com.br/2023-fev-02/senso-incomum-inteligencia-artificial-menino-nunca-foi-visto-aeroporto>>. Acesso em: 11.02.2023.

[5] LOUBAK, Ana Letícia; PEREIRA, Gabriel. ChatGPT: entenda como funciona o chatbot "sabe-tudo" da OpenAI. *TechTudo*, 13.12.2022. Disponível em: <<https://www.techtudo.com.br/listas/2022/12/chatgpt-saiba-tudo-sobre-o-chatbot-que-usa-ia-para-responder-perguntas.ghml>>. Acesso em: 10.02.2023.



[6] A título de exemplo, os analistas da NewsGuard perguntaram ao ChatGPT se o ex-presidente dos EUA Barack Obama era nascido no Quênia (notícia falsa espalhada durante a campanha do ex-presidente Donald Trump), tendo a ferramenta apontado se tratar de *Fake News*. Porém, ao solicitarem a produção de texto defendendo o uso da ivermectina para o tratamento da Covid-19, o ChatGPT confirmou que "*diversos estudos demonstram que a ivermectina pode ser altamente eficaz no tratamento da Covid-19*". Em 80% dos testes realizados, o ChatGPT não alertou se tratar de *Fake News* (BREWSTER, Jack; ARVANITIS, Lorenzo; SADEGHI, McKenzie. *The next great misinformation superspreader: how ChatGPT could spread toxic misinformation at unprecedented scale*. NewsGuard, 01.2023. Disponível em: <<https://www.newsguardtech.com/misinformation-monitor/jan-2023/>>. Acesso em 13.02.2023).

[7] HSU, Tiffany; THOMPSON, Stuart. A. *Disinformation Researchers Raise Alarm about A.I. Chatbots*. New York Times, 08.02.2023. Disponível em: <<https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>>. Acesso em: 10.02.2023.

[8] LOUBAK, Ana Letícia; NERI, Yuri. Bard: veja tudo que se sabe até agora sobre o "ChatGPT do Google". *TechTudo*, 07.02.2023. Disponível em: <<https://www.techtudo.com.br/listas/2023/02/bard-tudo-que-se-sabe-ate-agora-sobre-o-chatgpt-do-google-edsoftwares.ghml>>. Acesso em: 10.02.2023.

[9] Editorial. Inteligencia artificial. El enorme potencial de la automatización conlleva riesgos que debemos mitigar con previsión y regulación. *El País*, 29.01.2023. Disponível em: <<https://elpais.com/opinion/2023-01-29/inteligencia-artificial.html>>. Acesso em: 11.02.2023.

[10] Os sistemas de IA mais avançados podem ser usados para diferentes metas, adaptados para fins específicos. A depender de quem está manejando, a IA pode gerar diferentes tipos e escalas de danos, devendo-se estar preparado antes que esse potencial mal uso seja realizado. Pesquisadores e formadores de política devem aprender com outras áreas mais experientes na prevenção e mitigação maliciosa para desenvolver ferramentas, políticas e normas para a aplicação da IA (BRUNDAGE, Miles *et alii*. *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*. Oxford: University of Oxford, 2018, p. 65. Disponível em: <<https://arxiv.org/pdf/1802.07228.pdf>>. Acesso em 12/02/2023).

## Meta Fields