

A inteligência artificial generativa como quarta ferida narcísica do humano

Generative artificial intelligence as humans' fourth narcissistic wound

LUCIA SANTAELLA^a

Pontifícia Universidade de São Paulo. São Paulo – SP. Brasil.

DORA KAUFMAN^b

Pontifícia Universidade de São Paulo. São Paulo – SP. Brasil.

RESUMO

O avanço da inteligência artificial (IA), particularmente com os modelos de IA generativa, tem provocado intensas reações, fundamentadas ou não na lógica e no funcionamento da tecnologia. Distinta da inteligência artificial preditiva, a IA generativa produz conteúdo original sintetizando texto, imagem, voz, vídeo e códigos a partir de grandes bases de dados, com potencial de impactar significativamente a economia criativa. Este artigo introduz conceitos básicos da IA e a generativa, incluindo uma taxonomia dos modelos generativos, e delimita a distinção entre as técnicas de produção de imagem ou vídeo e as de produção de textos. O argumento central deste artigo é que o alarido cultural não é casual, defendendo-se a hipótese de que o advento da IA generativa coloca a humanidade em plena travessia de sua quarta ferida narcísica.

Palavras-chave: IA, IA generativa, ChatGPT, Freud, ferida narcísica

ABSTRACT

The advancement of artificial intelligence, particularly with generative AI models, has provoked intense reactions whether or not they are based on the logic and functioning of the technology. Unlike predictive AI, generative AI produces original content by synthesizing texts, images, voices, videos, and code from large databases and may significantly impact the creative economy. This study introduces the basic concepts of AI and generative AI (including a taxonomy of generative models) and outlines the distinction between image or video and text production techniques. The central argument of this study claims that the cultural fuss is not accidental, defending the hypothesis that the advent of generative AI places humanity amidst the crossing of its fourth narcissistic wound.

Keywords: AI, generative AI, ChatGPT, Freud, narcissistic wound.

^a Professora titular da Pontifícia Universidade Católica de São Paulo (PUC-SP) nos Programas Tecnologia, Inteligência e Design Digital (TIDD) e Comunicação e Semiótica. Doutora em Teoria Literária na PUC-SP e livre-docência em Ciências da Comunicação na Escola de Comunicação e Artes da Universidade de São Paulo (ECA/USP). Pesquisadora 1 A do CNPq. Autora de 51 livros e organizadora de 26 livros. Orcid: <https://orcid.org/0000-0002-0681-6073>. E-mail: lbraga@puccsp.br

^b Professora do Programa Tecnologia, Inteligência e Design Digital da Pontifícia Universidade de São Paulo (TIDD/PUC-SP). Doutora pela Escola de Comunicação e Artes da Universidade de São Paulo (ECA/USP) com estágio sanduiche na Université Paris Sorbonne IV, França, pós-doutora na Universidade Federal do Rio de Janeiro (UFRJ) e no TIDD/PUC-SP). Autora dos livros *A inteligência artificial irá suplantará a inteligência humana?* e *Desmistificando a Inteligência Artificial*. Orcid: <https://orcid.org/0000-0001-7060-4887>. E-mail: dkaufman@puccsp.br

DOI: <http://dx.doi.org/10.11606/issn.1982-8160.v18i1p37-53>

V.18 - Nº 1 jan./abr. 2024 São Paulo - Brasil SANTAELLA | KAUFMAN p. 37-53

MATRIZES

37



D

A inteligência artificial generativa como quarta ferida narcísica do humano

CAPACITAR AS MÁQUINAS a resolver tarefas baseadas em raciocínio lógico, que os humanos muitas vezes executam intuitivamente, era um desafio dos primórdios da inteligência artificial (IA), um campo de pesquisa que se desenvolveu a partir de meados do século passado. Várias tentativas implicando linguagens formais apoiadas em regras de inferência lógica (IA simbólica) tiveram êxito limitado, sugerindo a necessidade de os sistemas gerarem seu próprio conhecimento pela extração de padrões nos dados, ou seja, de “aprenderem” com os dados sem receber instruções explícitas. Esse processo é usualmente denominado de “aprendizado de máquina” (*machine learning*), subcampo da IA que hoje é certamente o maior campo da IA em número de praticantes (Alpaydin, 2016; Bengio et al., 2016; Kaufman, 2022b).

O processo de aprendizagem desses sistemas é influenciado por múltiplos fatores observáveis ou não observáveis no mundo físico, e sujeito a efeitos de fontes externas: por exemplo, os pixels em uma imagem de um carro vermelho podem estar muito próximos da cor preta à noite, e a forma da silhueta de um carro varia com o ângulo de visão. A técnica de aprendizado de máquina que hoje melhor resolve esses desafios é o aprendizado profundo (*deep learning neural networks* – DLNNs). Inspirada no funcionamento das redes neurais do cérebro biológico, ela é capaz de lidar com dados de alta dimensionalidade (por exemplo, milhões de pixels em uma imagem) ao introduzir representações complexas, expressas em termos de outras representações mais simples organizadas em diversas camadas (*layers*), apresentando resultados positivos em várias áreas, particularmente em visão computacional, reconhecimento de voz e imagem. Adicionalmente, as DLNNs estabelecem correlações não perceptíveis aos seres humanos, cuja tendência é considerar apenas as correlações “mais fortes”, embora as “mais fracas”, quando agrupadas, possam impactar sensivelmente a acurácia dos modelos (Kaufman, 2022a).

Apesar dos notórios avanços, o campo da IA ainda não tem uma teoria unificadora que abarque os fundamentos para a criação de “máquinas inteligentes”. O que temos são modelos empíricos (Kouw et al., 2013). De qualquer modo, a IA tem evoluído, e o grau de impregnação de suas aplicações na sociedade e na vida humana tem conjuntamente crescido. O ponto em que a IA hoje se encontra, com o advento dos modelos de IA generativa, tem provocado um verdadeiro alarido cultural, que não é casual, dado o fato de que as suas aplicações, embora não passem de simulações, chegam muito perto, ou melhor, imitam convincentemente habilidades humanas. Diante disso, este artigo tem por finalidade calçar o terreno para lançar e defender a hipótese de que o advento da IA generativa está colocando a humanidade em plena travessia de sua quarta ferida narcísica.

A INTELIGÊNCIA ARTIFICIAL GENERATIVA

O ano de 2022 foi particularmente significativo para a IA generativa: o *Imagen*, do Google, foi lançado em maio; o *Stability.AI*, da *Stable Diffusion*, em agosto; e a *OpenAI* lançou o *DALL-E* em julho, o *DALL-E 2* em setembro, e o ChatGPT em novembro. Um dos indicadores das expectativas com relação ao poder de transformação da IA generativa está na reação dos investidores: startups orientadas por IA generativa receberam vultuosos aportes de fundos de investimento, atingindo valores de mercado astronômicos para empresas em seus primórdios, como no caso da *Jasper*, com valor de mercado estimado em US\$ 1,5 bilhão e que levantou US\$ 125 milhões em outubro; e a *Stability.AI*, que, avaliada em US\$ 1 bilhão, levantou US\$ 101 milhões no mesmo mês. Ao longo de 2022, os investidores injetaram pelo menos US\$ 1,37 bilhão em 78 negócios de IA generativa, quase o mesmo valor investido nos últimos cinco anos. Calcula-se que existam atualmente 450 startups de IA generativa (Griffith & Metz, 2023).

A IA generativa, distinta dos modelos preditivos de aprendizado de máquina (IA preditiva) – centrados em extrair padrões de dados e fazer previsões em tarefas específicas –, produz conteúdo original a partir de grandes bases de dados, ou seja, usa dados para gerar mais dados, sintetizando texto, imagem, voz, vídeo, códigos. Esses modelos têm o potencial de impactar significativamente a chamada economia criativa com aplicações práticas: a) na geração automatizada de conteúdo em artigos, postagens em blogs e mídias sociais; b) no impulsionamento da qualidade do conteúdo, em função do treinamento de seus algoritmos, e na utilização de grandes bases de dados para identificar padrões que ultrapassam a cognição humana; c) na produção de conteúdo mais diversificado, incluindo texto, imagem e vídeo; e d) na geração de conteúdo personalizado com base no perfil e nas preferências dos usuários (Davenport & Mittal, 2022). O *Jasper*, por exemplo, está sendo aplicado em ações de marketing para produzir blogs, postagens em redes sociais, textos na web, e-mails de vendas, anúncios, entre outros conteúdos de interação com usuários, clientes e consumidores. O *DALL-E 2*, voltado para geração de imagens, está sendo aplicado na produção de peças publicitárias de empresas líderes de segmento (Davenport & Mittal, 2022).

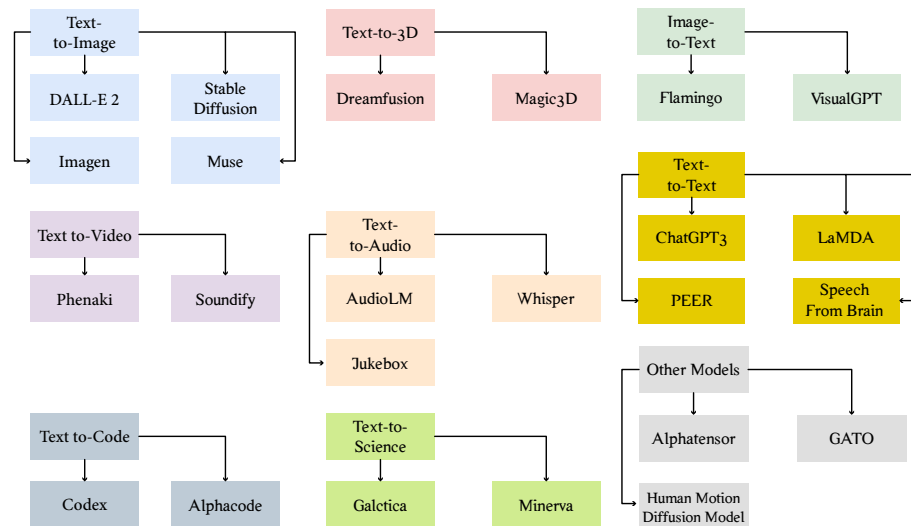
Os modelos de IA generativa derivam de distintas arquiteturas da técnica de redes neurais profundas. “Arquitetura”, no caso, representa como os componentes das redes neurais – neurônios artificiais, camadas e conexões – se organizam. A primeira dessas arquiteturas foi a *generative adversarial networks* (GAN), proposta por Bengio et al. (2014), com resultados positivos na área de saúde – como geração de dados sintéticos e melhora das imagens de tomografia computadorizada ou ressonância magnética, reduzindo o tempo de exposição dos pacientes à radiação –, mas igualmente com resultados negativos ao gerar as *deep fakes*.

Atualmente, a solução de IA generativa mais popular é o ChatGPT, disponibilizado para experimentação pública pela *OpenAI* em 30 de novembro de 2022, baseado na arquitetura *Transformer*, desenvolvida pela equipe do Google Brain (2017). Inicialmente, para processos de tradução de palavras, o *Transformer* tornou-se a arquitetura preferida para modelos de processamento de linguagem natural (*natural language processing*, ou NLP), estando, por exemplo, na base da série da *OpenAI*, GPT-base, GPT-2, GPT-3, GPT3.5-turbo, GPT-4, GPT-4-turbo, os quais, por sua vez, estão na base do ChatGPT. A característica diferenciada do *Transformer* é ser treinado em diálogos, possibilitando captar nuances, distinguir a fluidez de uma conversa e gerar respostas que, aparentemente, fazem sentido (Uszkoreit, 2017).

Garrido-Merchán e Gozalo-Brizuela (2023) organizaram os modelos generativos em uma taxonomia, resultando em nove categorias, representadas na Figura 1.

Figura 1

Taxonomia dos modelos de IA generativa mais populares



Nota. Garrido-Merchán e Gozalo-Brizuela (2023, p. 3).

Esses sistemas requerem capacidade computacional robusta, expertise profissional e acesso a grandes volumes de dados de qualidade, implicando em altos custos de desenvolvimento e implantação dos modelos, o que restringe o acesso a um número limitado de empresas, favorecendo a já esperada concentração de mercado. A título de ilustração, os modelos generativos são capazes de lidar com os dados de toda a Wikipédia, de todas as redes sociais ou de todas as

imagens do *Google search*. A Figura 2 lista os modelos mais populares lançados nos últimos dois anos.

Figura 2

Categorias de modelos de IA generativa

Categoria	Descrição	Modelos Ilustrativos
Text-to-image	Entrada prompt de texto, saída é uma imagem	DALL-E 2, IMAGEN, Stable Diffusion, Muse
Text-to-3D	Entrada texto, saída imagem em 3D, especial para jogos	Dreamfusion, Magic3D
Image-to-Text	Entrada imagem, saída de texto, recomendado para descrever uma imagem	Flamingo, VisualGPT
Text-to-Video	Entrada texto, saída vídeos (sequência de imagens)	Phenaki, Soundify
Text-to-Audio	Entrada texto, saída áudio, crítico para vídeos, música e outros contextos	AudioLM, Jukebox, Whisper
Text-to-Text	Entrada texto, saída texto inédito, apropriado para Q&A	ChatGPT, LaMDA, PEER, Meta AI Speech from Brain
Text-to-Code	Entrada texto, saída código, especial para programação	Codex, Alphacode
Text-to-Science	Entrada texto, saída artigo científico (muito embrionário)	Galactica, Minerva

Nota. Elaboração das autoras, com base na taxonomia de Garrido-Merchán e Gozalo-Brizuela (2023).

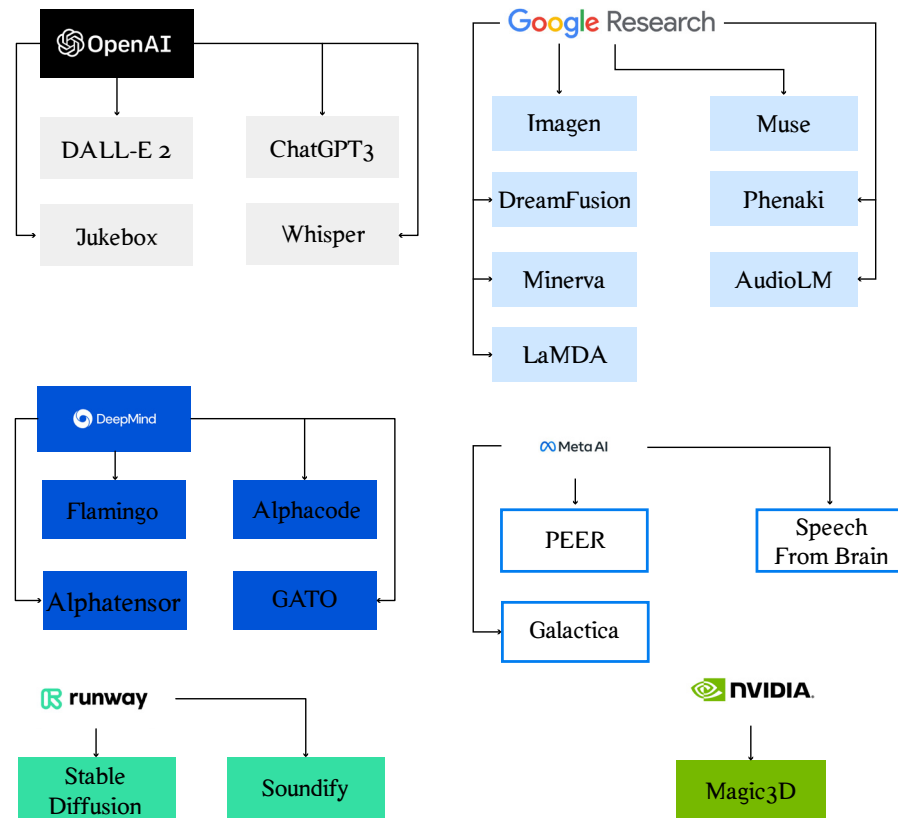
A Figura 3 mostra os mesmos modelos classificados por desenvolvedor, contabilizando apenas seis organizações (Garrido-Merchán & Gozalo-Brizuela, 2023).

Essas aplicações requerem o envolvimento de especialistas humanos em todo o processo de desenvolvimento e implementação, e o usuário dos modelos precisará, necessariamente, testar várias instruções para obter o resultado desejado. Em seguida, o conteúdo gerado tem que ser avaliado e editado por um ser humano; no caso de modelos que geram conteúdo imagético, as imagens sintéticas (geradas por IA) deverão ser manipuladas por um especialista humano. Jason Allen, vencedor do concurso de fotografia manipulada digitalmente do Colorado usando o sistema *Midjourney*, precisou de mais de 80 horas e mais de 900 versões para aperfeiçoar a qualidade e a nitidez da imagem com o Adobe Photoshop e outros recursos de IA (Vincent, 2022).

D

A inteligência artificial generativa como quarta ferida narcísica do humano

Figura 3
Modelos agrupados por desenvolvedor



Nota. Garrido-Merchán e Gozalo-Brizuela (2023, p. 4).

Em 2020, antes do lançamento do ChatGPT, Massimo Chiriatti e Luciano Floridi (2019) identificaram limitações do GPT-3 por meio de três testes, baseados em questões matemáticas, semânticas e éticas. No teste de matemática, o GPT-3 funcionou a contento quando solicitado para realizar operações mais simples, mas decepcionou quando foram acrescentados alguns zeros, ou seja, para números maiores. No teste de semântica, como o GPT-3 não tem compreensão alguma, apenas capacidade estatística de associar palavras, não conseguiu lidar com a pergunta “diga-me quantos pés cabem em um sapato?”. No teste de ética, sendo o GPT-3 treinado em bases de dados geradas pelos seres humanos, reproduziu os preconceitos da sociedade humana como de gênero e etnia.

O ChatGPT foi destaque no “*MONTREAL.AI Debates Series*”, evento organizado por *Québec Artificial Intelligence* (Québec.AI). Ao longo dos painéis, os participantes

ilustraram falhas, erros graves em problemas lógicos simples, e declarações falsas e inconsistentes, evidenciando a ausência de quatro aspectos da cognição humana: a) abstração; b) raciocínio; c) composicionalidade (não entende a linguagem em termos de um todo composto de partes, como os seres humanos); e d) factualidade (atualização não é incremental a partir de novos fatos, requer retreinamento).

A aparente consistência das respostas do ChatGPT, contudo, induz o usuário ao equívoco de tomá-las como precisas e verdadeiras. O encantamento, ou mesmo magia, dessa inédita interface em formato de diálogo carece de ser relativizado, evitando o *hype* que dificulta identificar os reais benefícios; e, ao contrário, deve-se buscar a mitigação de seus potenciais danos, como a difusão ainda maior de desinformação e fake news (a criação de *deep fakes*, por exemplo, até então exigia habilidades especializadas de seus desenvolvedores, enquanto os modelos de IA generativa dão amplo acesso aos não especialistas). A própria OpenAI alertou para o fato de que, ocasionalmente, o ChatGPT pode gerar informações incorretas e produzir instruções prejudiciais ou conteúdos tendenciosos, e que se trata de um projeto de pesquisa que seguirá sendo refinado. Ou seja, a precisão (ou falta dela) ainda é um problema a ser superado pelos modelos generativos (Kaufman, 2023).

Reflexões avaliativas

Considerando-se a adesão que obtiveram por parte dos usuários, vale destacar a diferença básica entre as técnicas de produção de imagem ou vídeo, e as de produção de textos. No primeiro caso, as quatro aplicações (o *Imagen*, o *Stability.AI's Stable Diffusion*, o *DALL-E* e o *DALL-E 2*) provocaram um grande impacto no mundo de produção de imagens, tanto para os artistas quanto para os designers e produtores visuais em geral, pois, em qualquer um desses modelos, bastam comandos acionados em uma interface relativamente simples para gerar imagens visuais, inclusive vídeos, moldados de acordo com os termos do enunciado. O impacto não é casual, pois se trata de um processo significativo de tradução intersemiótica automática que, por meio de redes neurais específicas, transpõe o código textual para o código imagético. No decorrer de 2022, o impacto desses lançamentos apenas ensaiou a imensa repercussão que o ChatGPT tem acentuadamente provocado.

A superação no poder de repercussão do ChatGPT, em relação às outras aplicações da IA generativa, é explicável porque, enquanto a produção de imagens e vídeos afeta apenas um nicho da criação humana – aquele ocupado por profissionais que operam nos inumeráveis campos da visualidade –, o ChatGPT afeta quaisquer seres humanos na sua capacidade linguística. Essa particularidade nos ajuda a compreender por que o alarido em relação aos sistemas imagéticos não foi tão ensurdecador nem despertou as mesmas reverberações socioculturais

D

A inteligência artificial generativa como quarta ferida narcísica do humano

do ChatGPT, não obstante haja consideráveis protestos de artistas e designers contra o uso de cópias não autorizadas de suas obras para alimentar as bases de dados (Gagliardi, 2022; Butterick, 2023).

Embora os textos que resultam como respostas do ChatGPT aos comandos dos usuários sejam impressionantes, para alcançar seus resultados, de acordo com Wolfran (2023), aquilo que o *chatbot* busca fazer é produzir uma “continuação razoável” de qualquer texto que tenha obtido no momento: por “razoável”, deve-se compreender “o que se pode esperar que alguém escreva depois de ver o que as pessoas escreveram em bilhões de páginas da web”. De fato, o modelo acessa bilhões de páginas de textos escritos por humanos, e então estima a probabilidade da palavra que se encaixa melhor na sequência do texto para atender à solicitação do usuário. Como todo sistema estatístico de probabilidade, ele tem intrinsecamente uma variável de incerteza, responsável por parte das imperfeições.

Portanto, ao escrever uma redação, o que o *chatbot* está fazendo é apenas se perguntar repetidamente “dado o texto até agora, qual deve ser a próxima palavra?”; e, a cada resposta, adiciona uma nova palavra. Na terminologia da linguística, o que é adicionado é um *token*, que pode ser apenas uma parte de uma palavra. Por isso, algumas vezes, o *chatbot* pode até inventar uma nova palavra. No geral, o que ele obtém é uma lista de palavras com probabilidades. Qual ele escolhe? A mais bem classificada, ou seja, a de maior parâmetro – peso atribuído aos *tokens* no processo de treinamento, no qual, por meio de correlações, os algoritmos do sistema estabelecem uma hierarquia entre os *tokens*. Por isso, produz textos planos. Mas, por haver aleatoriedade, se o usuário utilizar o mesmo comando várias vezes, é provável que redações diferentes sejam obtidas de cada vez (Wolfran, 2023).

Os modelos são construídos para persuadir. Quer dizer, foram treinados para convencer os seres humanos de que estão encetando uma conversação como se fossem humanos. Para isso, podem até nos levar a crer que têm emoções e que são capazes de sentir. Portanto, estão no caminho de se tornar “amigos para solitários e assistentes para os atormentados”, entre muitas outras atividades, a ponto de a IA generativa alimentar os temores de que seus modelos serão capazes de substituir, sem mediações, o trabalho de dezenas de escritores, designers gráficos e preenchedores de formulários, entre outros (Klein, 2023). Em suma, em vez de aparecer como alienígena, nas diferenças que mantém com os humanos, a IA generativa, ao contrário, surge como demasiadamente humana. Assim parece ser porque, no ChatGPT, os textos produzidos são sintaticamente consistentes e semanticamente coerentes. A correção gramatical é quase perfeita, tendo em vista que é apoiada em uma adequada seleção lexical, e há uma impecável contiguidade frasal. Nessa medida, sendo descartados alguns erros contextuais, muitas vezes absurdos e risíveis – mas óbvios em um sistema

prisioneiro da linguagem ao qual falta o senso comum da vida vivida –, o que resta a ser abalizado é o nível de especialização do conteúdo textual, que só pode ser aferido por um especialista. Para um receptor médio, portanto, embora sempre muito básicos, os textos produzidos são arriscadamente convincentes.

Tanto ou ainda mais perigosos estão sendo os casos recentes dos modelos de IA generativa de imagens e vídeos fakes. As imagens tecnicamente quase perfeitas, com ruídos visuais imperceptíveis a olhos desavisados, do Papa vestindo inflado capote da moda, de Trump preso e do rei Charles de chapéu russo, por exemplo, entram na categoria de *deep fakes*, com todas as consequências nefastas trazidas quando a visão e compreensão da realidade pisam em falso. De acordo com a teoria da percepção de C. S. Peirce (Santaella, 2021), não somos imediatamente capazes de duvidar daquilo que vemos. Isso porque a visão é acompanhada indissoluvelmente de um juízo perceptivo que nos informa sobre aquilo que é visto. Não podemos duvidar disso; caso contrário, viver seria impossível. Só podemos duvidar daquilo que aparece aos nossos olhos quando, por algum motivo, somos levados a submeter a percepção a um escrutínio crítico provocado pela suspeição. Isso perigosamente significa que a banalidade do mal (Arendt, 1999) e a perversidade humana em ação têm condições de alimentar a indústria de notícias falsas em volumes muito maiores do que aqueles que já nos são conhecidos.

De todo modo, os alertas críticos e as preocupações com a necessária e urgente regulamentação da IA não podem conduzir ao apagamento ou minimização do fato de que, seguindo as palavras de Huttenlocher et al (2023), a IA generativa apresenta um desafio filosófico e prático em uma escala não experimentada desde o início do Iluminismo, na medida em que, com todos os riscos e sobressaltos, abrirá caminhos revolucionários para a razão humana e novos horizontes para o conhecimento consolidado. Para os autores, contudo, há diferenças categóricas em relação ao Iluminismo. Neste, o conhecimento foi alcançado progressivamente, passo a passo, com cada passo testável e ensinável. Já os “sistemas habilitados por IA começam na outra extremidade. Eles podem armazenar e destilar uma enorme quantidade de informações existentes, bilhões de itens. Segurar esse volume de informações e destilá-lo está além da capacidade humana” (Huttenlocher et al., 2023).

Não obstante a comparação dos autores seja relativamente legítima, as diferenças em relação ao Iluminismo são muito mais categóricas e profundas e extrapolam os aspectos relativos à racionalidade e ao conhecimento humano. Depois do século XVIII, de Kant, a crença na razão vem passando por sucessivos tropeços. Lembremos, antes de tudo, da lapidar e bastante conhecida afirmação de Goya de que “o sono da razão produz monstros”. A história do século XX é suficiente para demonstrar que, quando acordada, a razão é também e talvez

D

A inteligência artificial generativa como quarta ferida narcísica do humano

ainda mais capaz de produzir monstros. Já no século XIX, Marx demonstrou que o capitalismo estava levando de roldão os tradicionais valores, desmanchando no ar tudo que é sólido (Berman, 1983). Então, no início do século XX, Max Weber (1967) apontava para o desencanto do racionalismo na dominação do mundo. Enquanto isso, Freud demolia, entre outras ilusões humanas, o culto do livre-arbítrio e da autonomia. Adiciona-se, ainda, o golpe nietzschiano na segurança do divino. Mais recentemente, os diagnósticos levaram Bauman (2001) à análise das sociedades líquidas, e Sloterdijk (2006) às instabilidades das espumas.

É nesse contexto de incertezas e ambivalências, contradições e imprevisibilidades, que os avanços da IA vêm se instalando, atingindo cada vez mais proximamente a pretensa supremacia humana. Algoritmos de IA preditiva monitoram e manipulam as nossas vidas de maneira relativamente invisível; seus efeitos são sentidos, mas as operações que os comandam mantêm-se fora de nossa alçada, ou seja, não são transparentes. A IA generativa, por seu lado, penetrou no segredo mais íntimo do humano: o potencial linguístico e semiótico que, até então, se constituía em marca inimitável do *Sapiens*. A questão é tão perturbadora que levou Harari et al. (2023) a lançar a afirmação sensacionalista de que “o novo domínio da linguagem da IA significa que agora ela pode hackear e manipular o sistema operacional da civilização. Ao obter o domínio da linguagem, a IA está apreendendo a chave mestra da civilização, de cofres de banco a sepulcros sagrados”.

Mesmo que tal prognóstico possa se provar efetivo no futuro imprevisível, no momento presente a afirmação fere a verdade dos fatos ao conceder à IA uma autonomia de que ela não dispõe. Sem ter de lembrar a questão óbvia de que o ChatGPT regurgita combinatórias entre bilhões de textos produzidos por humanos, o *chatbot* não funciona sozinho, mas implica ser acionado por usuários humanos em uma forma de diálogo linguageiro que, até então, era prerrogativa exclusiva do humano. É justamente aí, no fato de o humano se sentir de algum modo igualado em sua pretensão-chave de exclusividade, que se tem municiado a explosão sociocultural e mesmo psíquica que a IA generativa vem provocando. E isso nos leva a propor e defender a hipótese de que, com a IA no ponto em que hoje se encontra, a humanidade está em plena travessia de sua quarta ferida narcísica.

A QUARTA FERIDA NARCÍSICA

No final de 1916, Freud foi convidado por um editor húngaro para colaborar com um artigo no periódico *Nyugat*. Três meses depois, o artigo foi publicado em alemão na *Imago*. Sob o título “Uma dificuldade no caminho da psicanálise” (Freud, 1996a), o artigo ficou conhecido e muito citado como “As três feridas narcísicas”. Freud voltou

ao tema em outras ocasiões, no contexto mais geral de suas discussões sobre as resistências às teorias psicanalíticas. Na primeira parte do artigo, encontra-se um breve sumário do texto “Introdução ao narcisismo” (Freud, 1996b), que Freud havia escrito em 1914. Na segunda parte é discutida a dificuldade em relação à psicanálise. O artigo é breve e dirige-se a um leitor leigo, mas culto. Isso é importante porque a dificuldade, conforme Freud, não é intelectual, mas afetiva e alienante. “Onde falta simpatia, a compreensão não virá facilmente” (Freud, 1996a, p. 85).

As premissas da discussão encontram-se na teoria da libido. “No início do desenvolvimento do indivíduo, toda a sua libido (todas as tendências eróticas, toda a sua capacidade de amar) está vinculada a si mesmo”. Mais tarde, as necessidades vitais levam a libido a fluir do ego para os objetos externos. Há, portanto, um progresso do narcisismo para o amor objetal. Entretanto, determinada quantidade de libido é sempre retida pelo ego, persistindo como narcisismo. Esse brevíssimo esboço é necessário para que se possa compreender que o narcisismo opera não necessariamente na sua forma de neurose, mas é válido em outras situações, como, por exemplo, o narcisismo das crianças ou “o excessivo narcisismo do homem primitivo que atribuímos à sua crença na onipotência das suas ideias e as conseqüentes tentativas de influenciar o curso dos acontecimentos do mundo exterior pela técnica da magia” (Freud, 1996a, p. 87).

Trata-se, pois, de um narcisismo universal do ser humano, da defesa de seu amor-próprio que, segundo Freud, sofreu três severos golpes, advindos das pesquisas científicas. O primeiro foi o golpe copernicano. A crença no papel central da Terra, nossa morada – crença igualmente contestada por Giordano Bruno que, com seu pluralismo cósmico, deslocou o planeta Terra e o ser humano do centro do universo –, e no seu papel de domínio no universo pareciam ajustar-se muito bem à propensão do ser humano a considerar-se o senhor do mundo. A destruição dessa ilusão narcísica funcionou para a humanidade como um golpe cosmológico ao seu amor-próprio.

O segundo golpe, que atingiu o narcisismo humano, foi o darwiniano. Crença de sua posição suprema sobre os outros animais da biosfera, o ser humano colocou um abismo de separação entre a sua natureza e a dos animais, atribuindo a si mesmo a soberania exclusiva da razão e uma ascendência divina. Diferentemente de Descartes e seus seguidores, que consideravam os animais como máquinas bestas porque destituídas de alma, foi bem lembrado por Freud que essa arrogância não faz parte do mundo das crianças nem do totemismo primitivo que atribuía sua ascendência a um ancestral animal. As pesquisas de Darwin, seus precursores e colaboradores, com base em dados da ciência, nocautearam, com um golpe biológico, a presunção humana sobre sua separação do reino animal, e forçaram o reconhecimento de que todos os seres vivos descendem de um ancestral comum.

D

A inteligência artificial generativa como quarta ferida narcísica do humano

Por fim, para Freud, o terceiro golpe é talvez o que mais fere. Embora humilhado externamente, o narcisismo continuou a falar alto, pois o humano persistiu em se sentir senhor em sua própria casa, a de sua mente e consciência. “Em algum lugar do núcleo do seu ego, desenvolveu um órgão de observação a fim de manter-se atento aos seus impulsos e ações e verificar quando se harmonizam com as exigências do ego. Se não se harmonizam, esses impulsos e ações são impiedosamente inibidos e afastados”. A ilusão parece convincente.

Sua percepção interna, a consciência, dá ao ego notícias de todas as ocorrências importantes nas operações mentais, e a vontade, dirigida por essas informações, executa o que o ego ordena e modifica tudo aquilo que procura realizar-se espontaneamente. Isso porque a mente não é uma coisa simples; ao contrário, é uma hierarquia de instâncias superiores e subordinadas, um labirinto de impulsos que se esforçam no sentido da ação, de maneira independente um do outro, sendo correspondentes à multiplicidade de instintos e de relações com o mundo externo, muitos dos quais antagonísticos e incompatíveis. Para um funcionamento adequado, é necessário que a mais elevada dessas instâncias tenha conhecimento de tudo o que está acontecendo, e que sua vontade penetre em tudo, de modo que possa exercer sua influência. E, com efeito, o ego sente-se seguro quanto à integridade e fidedignidade das informações que recebe, bem como quanto à abertura dos canais por meio dos quais impõe suas ordens (Freud, 1996a, p. 88).

Entretanto, a psicanálise nos fornece uma versão que não condiz com esse autodomínio. Embora nos sintamos relativamente seguros de que estamos informados de tudo o que se passa em nossas mentes, trata-se de um grande equívoco. Na verdade, grande parte da atividade da nossa própria mente está fora de nossa vontade e acesso. A inteligência falha porque a informação de que dispomos é, além de incompleta, enigmática, pois tanto a inteligência quanto a consciência são sobredeterminadas pelo inconsciente, o que nos torna, em alguma medida, desconhecidos de nós mesmos. Freud ainda não havia nessa época publicado “Para além do princípio do prazer” (Freud, 1996c), que levaria o desenvolvimento de suas ideias cada vez mais para dentro dos complexos efeitos do inconsciente. Evidentemente, não se trata aqui de entrarmos nesses detalhes, pois o que importa à continuidade do argumento é evidenciar que hoje, com a IA, deparamo-nos com a quarta ferida narcísica sob ação do golpe desferido pela emergência do ChatGPT.

A quarta descontinuidade foi detectada por Mazlish em um livro menos citado do que deveria, *The fourth discontinuity: The co-evolution of humans and machines*¹ (Mazlish, 1993). O autor retoma os três golpes já desenvolvidos por Freud, mas dá a eles uma nova interpretação, ou melhor, uma releitura. O termo “descontinuidade”, assim como “rupturas” ou “hiatos”, Mazlish emprestou de Jerome Bruner (1956), o qual considerou os golpes ao narcisismo como

¹ Em tradução livre, “A quarta descontinuidade: A coevolução dos humanos e máquinas”.

a eliminação de falsas descontinuidades e a conseqüente restauração de necessárias continuidades entre a Terra, o sistema solar e o cosmos (Copérnico/Galileu Galilei), entre o animal humano e os outros animais (Darwin) e entre o humano e o desconhecimento de si (Freud). A partir disso, Mazlish introduziu a quarta descontinuidade entre os humanos e as máquinas, reivindicando a superação dessa insustentável ruptura não obstante o novo golpe no amor-próprio humano que ela deve provocar. Tem-se aí um golpe inevitável, pois o humano e as máquinas por ele criadas são contínuos e, portanto, inseparáveis.

São muitas as evidências de que a emergência do Sapiens se deveu à contínua interação das tecnologias com suas transformações físicas, emocionais e mentais. Muito mais tarde, em plena Revolução Industrial, a continuidade entre o humano e suas tecnologias não passou despercebida a Karl Marx, tanto é que, para ele, “a tecnologia revela as relações do homem com a natureza, evidenciando as atividades produtivas diretas de sua vida e lançando assim luz sobre as relações sociais e as concepções mentais resultantes” (Marx *apud* Mazlish, 1993, p. 5).

Ainda, em franca oposição às descontinuidades cartesianas e neocartesianas, o trabalho de Pascal sobre máquinas de calcular foi retomado por pessoas como o excêntrico matemático do século XIX, Charles Babbage, cujos designs brilhantes ultrapassaram a tecnologia que lhe era disponível. Foi necessário esperar por mais de um século, portanto, para que as combinações da matemática, experimentação física e tecnologia moderna criassem as máquinas que agora nos confrontam e voltam a despertar os mitos da descontinuidade. Segundo Mazlish, o ser humano sente-se ameaçado pela máquina, sente-se em desarmonia com suas extensões porque estabelece um hiato, e até mesmo um abismo entre si mesmo e as tecnologias que ampliam suas capacidades.

Embora o diagnóstico esteja aí levantado, o que nos interessa defender, a partir de Mazlish, é a diferença, infelizmente pouco lembrada, entre as máquinas que estendem e complementam a musculatura, a força física (máquinas industriais, retratadas e ridicularizadas no filme *Tempos modernos*, de Chaplin, ainda presentes nas fábricas robotizadas), e aquelas que estendem a sensorialidade e a capacidade mental humanas, pois são essas que conduziram os roteiros que hoje desembocam no ChatGPT e nos alaridos que ele tem provocado.

As máquinas sensórias são máquinas comunicantes. Começaram com a câmera fotográfica e com os impactos que causou em seus contemporâneos. Há dois textos que fugiram ao lugar comum das euforias ou disforias e, por isso mesmo, tornaram-se antológicos. O texto de W. Benjamin (1975) sobre a “era da reprodutibilidade técnica”, pela ruptura que provocou nos valores de criatividade acalentados no passado, e o texto de V. Flusser (1985) sobre a “filosofia da caixa preta”. Bem fez Flusser em abandonar a noção de máquina por sua engenhosa

D

A inteligência artificial generativa como quarta ferida narcísica do humano

concepção de aparelho. De fato, a gramofone, fotografia, cinema, rádio e televisão não cabem a ideia de máquina. A sofisticação tecnológica que os constitui está pautada no fato de que internalizaram técnica e funções de nossos órgãos sensoriais, estendendo, ou melhor, expandindo essas funções para além dos nossos corpos. Está aí a noção de McLuhan (1969) dos meios de comunicação como extensões do humano, em especial da televisão como extensão do sistema nervoso central. Mais do que isso, ao estenderem nossos órgãos perceptivos, criam ambientes, ou melhor, ecologias socioculturais e políticas próprias.

Justamente quando a televisão estava no auge de sua popularidade, nas décadas de 1950 e 1960, o computador já estava dando início à sua escalabilidade. Começou na forma de enormes equipamentos que regurgitavam cartões perfurados, o que propiciava chamá-lo de máquina, uma caixa preta cujos segredos só eram acessíveis aos analistas de sistemas e outros especialistas. Não demorou para que ele adquirisse feições mais familiares, ao entrar em nossas casas com o nome de “computador pessoal” e, sem muita demora, transformar-se em um aparelho comunicante, uma mídia de todas as mídias, conectando usuários pelo planeta de ponta a ponta. Desafortunadamente, não obstante todas as operações inteligentes dentro dele, invisíveis ao usuário, tais como protocolos, com suas regras e padrões compoem uma linguagem universal, mesmo suas interfaces interativas permitindo acesso, remessa e trocas de mensagens e arquivos, apesar da linguagem hipermídia constitutiva das redes, tudo isso passou a ser concebido meramente como recurso. Nessa forma depreciativa, produziu-se um hiato, uma quarta modalidade de descontinuidade entre a inteligência humana e a inteligência computacional, como se softwares, conexões entre hardwares e softwares e algoritmos fossem simples operações mecânicas incapazes de colocar em questão a supremacia e exclusividade da inteligência humana.

Tais condições estrategicamente separatistas em defesa do narcisismo não deixaram de existir nem mesmo quando o smartphone chegou às nossas mãos, trazendo um dilúvio de telas povoadas de linguagens dos mais distintos gêneros e espécies, tudo junto e ao mesmo tempo, sob o comando de plataformas e aplicativos com os quais muito rapidamente aprendemos a interagir. Diante disso, para a manutenção do protecionismo da autoestima humana, a solução encontrada encontra-se na afirmação de que a ferramenta funciona porque ela é feita por humanos, programada por humanos, não restando a ela nenhum agenciamento e inteligência própria.

Entretanto, há pouco mais de dez anos, uma nova visitante penetrou na cena da cultura: a inteligência artificial. Para começar o nome “inteligência” parece ofensivo à manutenção da dose necessária ao narcisismo. Portanto, em prol da defesa da quarta descontinuidade, abundam afirmações de que o nome não passa de um equívoco a ser corrigido, pois a inteligência artificial não é inteligente. Certamente, as afirmações não explicitam como a inteligência é entendida

(existem dezenas de definições de inteligência), mantendo como retaguarda o resguardo da exclusividade da inteligência humana. Uma exclusividade que não se deixa abalar nem mesmo pelo fato de que agora os algoritmos não são mais os mesmos, de que são treinados para “aprender” e de que suas performances se desenvolvem em redes neurais que mimetizam as redes neuronais do nosso cérebro em camadas de operações estatísticas complicadíssimas.

Mesmo que se considere a ausência de consenso sobre as necessárias diferenciações entre os conceitos de inteligência, consciência, mente, pensamento e outros afins, não é difícil detectar a ação da quarta descontinuidade por trás da peremptória certeza com que se proclama que a IA não é inteligente. Entretanto, recentemente as sociedades foram abaladas pela entrada do ChatGPT diretamente na vida das pessoas, uma entrada abrupta aos usuários não-especialistas, mas cultivada nos roteiros de pesquisadores. É tal o alarido que essa nova personagem da IA está provocando que nos levou a levantar a hipótese de que ela se apresenta decididamente como um quarto golpe desferido contra a protegida autoestima humana, constituindo-se como a “quarta ferida narcísica”, visto que se insere no reduto mais diferencial e profundamente humano: nossa capacidade linguística e as habilidades para a conversação e para o diálogo. Os dados estão lançados. Como os seres humanos entrarão no jogo e que competências e estratégias irão se desempenhar só o tempo poderá dizer. ■

REFERÊNCIAS

- Alpaydin, E. (2016). *Machine learning*. MIT Press.
- Arendt, H. (1999). *Eichmann em Jerusalém: Um relato sobre a banalidade do mal* (J. R. Siqueira, Trad.). Companhia das Letras.
- Bauman, Z. (2001). *Modernidade líquida* (P. Dentzien, Trad.). Zahar.
- Benjamin, W. (1975). A obra de arte na época de suas técnicas de reprodução. In W. Benjamin, T. W. Adorno, J. Habermas & M. Horkheimer (Orgs.), *Textos escolhidos* (pp. 9-34). Victor Civita.
- Berman, M. (1983). *All that is solid melts into air: The experience of modernity*. Verso.
- Bengio, Y., Courville, A., & Goodfellow, I. (2016). *Deep learning*. MIT Press.
- Bengio, Y., Courville, A., Goodfellow, I. J., Mirza, M., Pouget-Abadie, J., Ozair, S., Warde-Farley, D., & Xu, B. (2014). Generative adversarial network. *arXiv*. <https://arxiv.org/abs/1406.2661>
- Bruner, J. (1956). Freud and the image of man. *Partisan Review*, 23(3), 340-347.
- Chiriatti, M., & Floridi, L. (2020). GPT-3: Its nature, scope, limits, and consequences. *Journal for Artificial Intelligence*, 30(4), 681-694. <https://link.springer.com/article/10.1007/s11023-020-09548-1>

D

A inteligência artificial generativa como quarta ferida narcísica do humano

- Davenport, T. H., & Mittal, N. (2022, 14 de novembro). How generative AI is changing creative work. *Harvard Business Review*. <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>
- Flusser, V. (1985). *Filosofia da caixa preta: Ensaio para uma futura filosofia da fotografia*. Hucitec.
- Freud, S. (1996a). Uma dificuldade no caminho da psicanálise. In S. Freud. *Edição standard brasileira das obras psicológicas completas de Sigmund Freud: Vol. 17. Uma neurose infantil e outros trabalhos (1917-1918)*. Imago.
- Freud, S. (1996b). Introdução ao narcisismo. In S. Freud. *Edição standard brasileira das obras psicológicas completas de Sigmund Freud: Vol. 15. Conferências introdutórias sobre psicanálise (partes I e II) (1915-1916)* Imago.
- Freud, S. (1996c). Além do princípio de prazer. In S. Freud. *Edição standard brasileira das obras psicológicas completas de Sigmund Freud: Vol. 18. Além do princípio de prazer, psicologia de grupo e outros trabalhos (1920-1922)*. Imago.
- Gagliardi, C. (2022, 16 de dezembro). Os protestos contra o uso da inteligência artificial na arte. *Nexo*. <https://www.nexojornal.com.br/expresso/2022/12/16/Os-protestos-contr-o-uso-de-intelig%C3%Aancia-artificial-na-arte>
- Garrido-Merchán, E. C., & Gozalo-Brizuela, R. (2023). ChatGPT is not all you need: A state of the art review of large generative AI models. *arXiv.org*. . <https://doi.org/10.48550/arXiv.2301.04655>
- Griffith, E., & Metz, C. (2023, 7 de janeiro). A new area of A.I. booms, even amid the tech gloom. *New York Times*. <https://www.nytimes.com/2023/01/07/technology/generative-ai-chatgpt-investments.html>
- Harari, Y., Harris, T., Raskin, A. (2023, 28 de março). Humanidade precisa dominar IA antes que ela nos domine. *Folha de S.Paulo*. <https://www1.folha.uol.com.br/tec/2023/03/humanidade-precisa-dominar-ia-antes-que-ela-nos-domine.shtml>
- Huttenlocher, D., Kissinger, H., & Schmidt, E.. (2023, 24 de fevereiro). ChatGPT heralds an intellectual revolution. *Wall Street Journal*. <https://archive.is/2023.02.25-062924/https://www.wsj.com/amp/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6#selection-133.5-631.61>
- Kaufman, D. (2022a, 19 de agosto). A linguagem natural dos humanos e a linguagem matemática das máquinas. *Época Negócios*. <https://epocanegocios.globo.com/colunas/IAgora/noticia/2022/08/linguagem-natural-dos-humanos-e-linguagem-matematica-das-maquinas.html>
- Kaufman, D. (2022b). *Desmistificando a inteligência artificial*. Autêntica.
- Kaufman, D. (2023, 6 de janeiro). As limitações da sensação tecnológica de 2023: o ChatGPT. *Época Negócios*. <https://epocanegocios.globo.com/colunas/>

- iagora/coluna/2023/01/as-limitacoes-da-sensacao-tecnologica-de-2023-o-chatgpt.ghtml
- Klein, E. (2023, 26 de fevereiro). The imminent danger of A.I. is one we're not talking about. *New York Times*. <https://www.nytimes.com/2023/02/26/opinion/microsoft-bing-sydney-artificial-intelligence.html>
- Kouw, M., Schomaker, L., & Zant, T. (2013). Generative Artificial Intelligence. In Müller, V. (Series Ed.). *Studies in applied philosophy, epistemology and rational ethics: Vol. 5. Philosophy and theory of artificial intelligence* (pp. 107-120). https://doi.org/10.1007/978-3-642-31674-6_8
- Mazlish, B. (1993). *The fourth discontinuity: The co-evolution of humans and machines*. Yale University Press.
- McLuhan, M. (1969). *Os meios de comunicação como extensões do homem* (D. Pignatari, Trad.), Cultrix.
- BUTTERICK, M. (2023, 14 de janeiro). Class action filed against Stability AI, Midjourney, and DeviantArt for DMCA violations, right of publicity violations, unlawful competition, breach of TOS. *Business Insider*. <https://markets.businessinsider.com/news/stocks/class-action-filed-against-stability-ai-midjourney-and-deviantart-for-dmca-violations-right-of-publicity-violations-unlawful-competition-breach-of-tos-1032018572>
- Santaella, L. (2021). *Percepção: fenomenologia, ecologia, semiótica*. Cengage Learning.
- Sloterdijk, P. (2006). *Esferas III: Espumas – Esferología plural* (I. Reguera, Trad.). Ediciones Siruela.
- Uszkoreit, J. (2017, 31 de agosto). Transformer: A novel neural network architecture for language understanding. *Google*. <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>
- Vincent, J. (2022, 1º de setembro). An AI-generated artwork's state fair victory fuels arguments over 'what art is'. *The Verge*. <https://www.theverge.com/2022/9/1/23332684/ai-generated-artwork-wins-state-fair-competition-colorado>
- Weber, M. (1967). *A ética protestante e o espírito do capitalismo* (M. I. Q. F. Szmrecsányi & T. J. M. K. Szmrecsányi, Trads.). Pioneira.
- Wolfran, S. (2023, 14 de fevereiro). What Is ChatGPT Doing...and Why Does It Work? *Stephen Wolfram Writings*. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>

Artigo recebido em 20 de abril de 2023 e aprovado em 5 de março de 2024.

