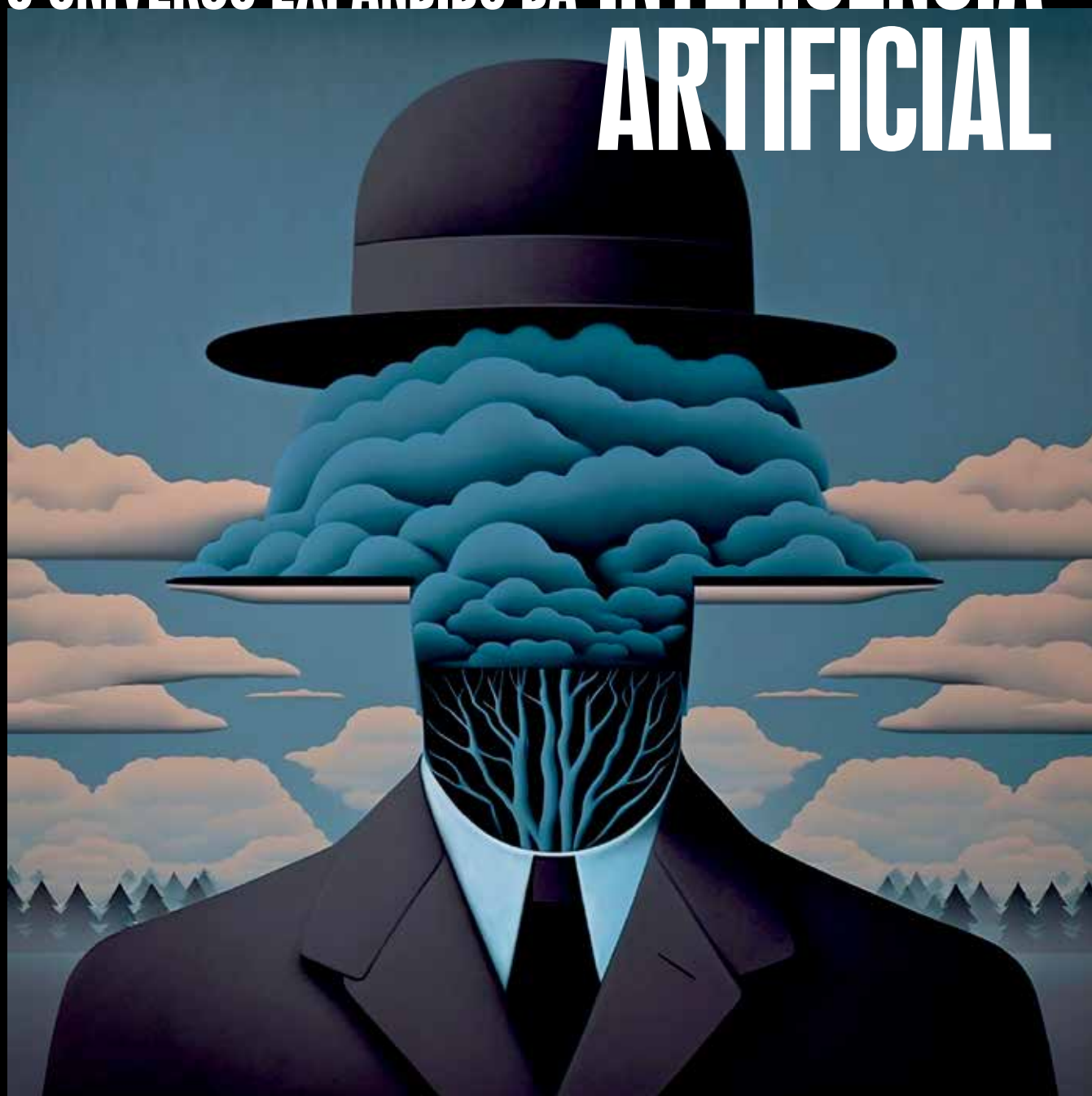


CAPA

# O UNIVERSO EXPANDIDO DA INTELIGÊNCIA ARTIFICIAL



# ChatGPT reacende debate sobre o potencial criativo de sistemas de linguagem natural e as implicações éticas relacionadas ao seu uso

Rodrigo de Oliveira Andrade

**H**á anos cientistas em empresas e universidades criam e aperfeiçoam softwares de inteligência artificial (IA), mas poucas vezes essa tecnologia galvanizou a atenção da sociedade como em fins de 2022, com o ChatGPT, sistema capaz de criar textos realistas e articulados a partir de perguntas e comandos dos usuários e simular conversas com um ser humano. Lançada em novembro pela OpenAI, startup com sede na Califórnia, nos Estados Unidos, a ferramenta foi a protagonista de um experimento coletivo de alcance planetário. Dois meses após seu lançamento, já era usada por mais de 100 milhões de pessoas, desencadeando uma explosão de experiências de escrita, algumas divertidas, outras preocupantes. O modelo se mostrou capaz de compor músicas, escrever poemas, códigos de programação e até textos jornalísticos – para inquietação de repórteres como eu –, reacendendo a discussão: qual é o risco de a inteligência artificial, mimetizando a nossa capacidade criativa e levando-a a novos patamares, substituir os seres humanos em atividades intelectuais? Até aqui, esse impacto parecia restrito a tarefas de caráter repetitivo.

Ainda é cedo para saber o quanto essa tecnologia transformará a vida das pessoas. É certo que ela abre uma fronteira de novas aplicações, prometendo reinventar desde mecanismos de busca on-line a assistentes de voz, como Alexa e Siri, e permitindo que pessoas conversem com computadores e outros dispositivos eletrônicos como se estivessem falando com humanos. “Há todo um mundo novo de comunicação que essa tecnologia vai inaugurar”, disse ao serviço de notícias *Stanford News* Jeff Hancock, professor da Universidade Stanford, nos Estados Unidos, que estuda como a comunicação mediada por sistemas de inteligência artificial afeta relacionamentos interpessoais.

O funcionamento do ChatGPT se baseia em algoritmos de redes neurais profundas, modelos que se inspiram na organização do sistema nervoso humano e se apoiam em aprendizado

de máquina, um campo de estudo que permite extrair padrões de grandes volumes de dados e fazer previsões a partir deles. Esse tipo de sistema funciona com base em unidades de processamento interconectadas em várias camadas, da mesma forma que os neurônios se conectam por sinapses. Nos últimos anos, empresas como Google e OpenAI passaram a se dedicar ao desenvolvimento de um tipo específico de rede neural artificial, os grandes modelos de linguagem (LLM), os quais conseguem analisar enormes quantidades de dados de texto, identificar bilhões de padrões sobre como as pessoas conectam palavras, números e símbolos, e, a partir disso, aprender a gerar textos novos. O desenvolvimento dessas ferramentas levou à criação do que os cientistas chamam de sistemas de IA generativa.

Desde 2018, o Google usa essa tecnologia em seus sites de pesquisa e de tradução a partir de um software chamado Bert, que parte de modelos probabilísticos para prever as próximas palavras de um texto com base no trecho anterior – é dessa forma que o Google corrige a ortografia de palavras e termina frases de usuários que usam seu mecanismo de busca.

Em fins de 2020, a OpenAI deu um salto e lançou o GPT-3, programa de processamento de linguagem capaz de analisar e processar dados de texto em uma escala sem precedentes. O GPT-3 passou meses em treinamento, escrutinando centenas de gigabytes de informações na internet, incluindo livros, artigos científicos, páginas da Wikipédia, notícias e registros de bate-papo on-line, a partir dos quais identificou 175 bilhões de parâmetros, isto é, representações matemáticas de padrões de texto – para se ter ideia, o Bert trabalhava com “apenas” 110 milhões de parâmetros. “Esses padrões equivalem a um mapa da linguagem humana, uma descrição matemática do modo como escrevemos textos e códigos de programação”, esclarece o cientista da computação Fernando Santos Osório, do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP).

Tais mapas permitem que o GPT-3 escreva sobre qualquer assunto em qualquer estilo, mesmo

Ilustração produzida por software de IA a partir do comando “redes neurais profundas de inteligência artificial, estilo René Magritte”

+ Novo chat

## PERGUNTAS FREQUENTES



**O que é o ChatGPT?**



É um software de IA generativa que produz textos a partir de perguntas e comandos dos usuários



**Como ele funciona?**



Com base em redes neurais profundas chamadas “grandes modelos de linguagem”



**O que elas fazem?**



Analizam enormes quantidades de dados de texto e identificam bilhões de padrões sobre como os seres humanos conectam palavras, números e símbolos, aprendendo a gerar textos novos em linguagem natural



**Qual a quantidade de dados analisados pelo ChatGPT?**



Estima-se que o ChatGPT tenha analisado 570 gigabytes de dados de textos na internet (ou 300 bilhões de palavras), a partir dos quais identificou 175 bilhões de parâmetros



**O que são parâmetros?**

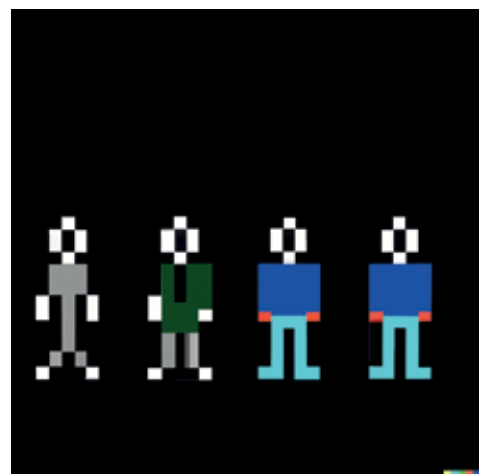


Representações matemáticas de padrões de texto, uma espécie de descrição matemática do modo como seres humanos escrevem textos e códigos de programação

que a solicitação pareça surreal, como um poema sobre a conquista do Palmeiras na Copa Rio de 1951 segundo as regras de versificação da *Odisseia*, de Homero, ou roteiros de comédias românticas em universos alternativos. Também consegue explicar conceitos de mecânica quântica, em prosa clara e bem pontuada, criando a impressão de que estamos mesmo falando com um ser humano, e não interagindo com um software. A OpenAI aplicou a mesma lógica em outro tipo de tecnologia de IA generativa, o DALL-E, treinado com base em dezenas de milhares de fotos digitais, a partir das quais consegue gerar novas imagens, como algumas das que ilustram esta reportagem.

**O** ChatGPT utiliza uma versão mais avançada do GPT-3, o GPT-3.5, em uma interface de conversação amigável e fácil de usar. Diferentemente de seu antecessor, que foi testado por um pequeno grupo de usuários, a ferramenta foi disponibilizada gratuitamente para o público, amplificando a percepção sobre seu potencial disruptivo. Em termos das técnicas que utiliza, o ChatGPT não é particularmente inovador. “Tampouco é revolucionário, embora seja percebido assim pelo público, possivelmente por oferecer uma solução interativa inédita”, comenta Dora Kaufman, professora da Faculdade de Ciências e Tecnologia da Pontifícia Universidade Católica de São Paulo (PUC-SP).

Ao permitir que milhares de pessoas pudessem experimentar sua tecnologia, a OpenAI despertou ao mesmo tempo o fascínio do público e o apetite de investidores. O principal deles, a Microsoft. A multinacional, que já havia investido US\$ 1 bilhão na startup em 2019, aportou mais US\$ 10 bilhões em janeiro deste ano na expectativa de poder incorporar a ferramenta a vários de seus serviços. O motor de busca Bing, por exemplo, funciona des-



de o início de fevereiro com base em um sistema de IA mais poderoso que o ChatGPT. O modelo de linguagem foi desenvolvido em parceria com OpenAI e em breve deverá ser incorporado ao navegador Microsoft Edge.

Os concorrentes logo reagiram. Dias após o anúncio da Microsoft, a Alphabet, dona do Google, lançou o Bard, seu próprio sistema de IA generativa, que deverá igualmente ser incorporado ao seu sistema de busca. A empresa adquiriu ainda uma participação de US\$ 300 milhões na Anthropic, startup fundada por ex-funcionários da OpenAI e responsável por um software parecido com o ChatGPT, o Claude. Na China, as ações da multinacional de tecnologia Baidu subiram 15% após o anúncio, em fevereiro, de que lançará um robô de IA, batizado de Ernie Bot (versão em inglês) e de Wenxin Yiyan (em mandarim).

Nenhuma outra tecnologia atraiu tanto dinheiro nos últimos anos como a IA generativa, “possivelmente porque essa é uma área que costuma dar retornos rápidos sobre o capital investido”, comenta o economista Alexandre Chiavegatto Filho, diretor do Laboratório de Big Data e Análise Preditiva em Saúde da Faculdade de Saúde Pública da USP, que estuda o uso de métodos predi-

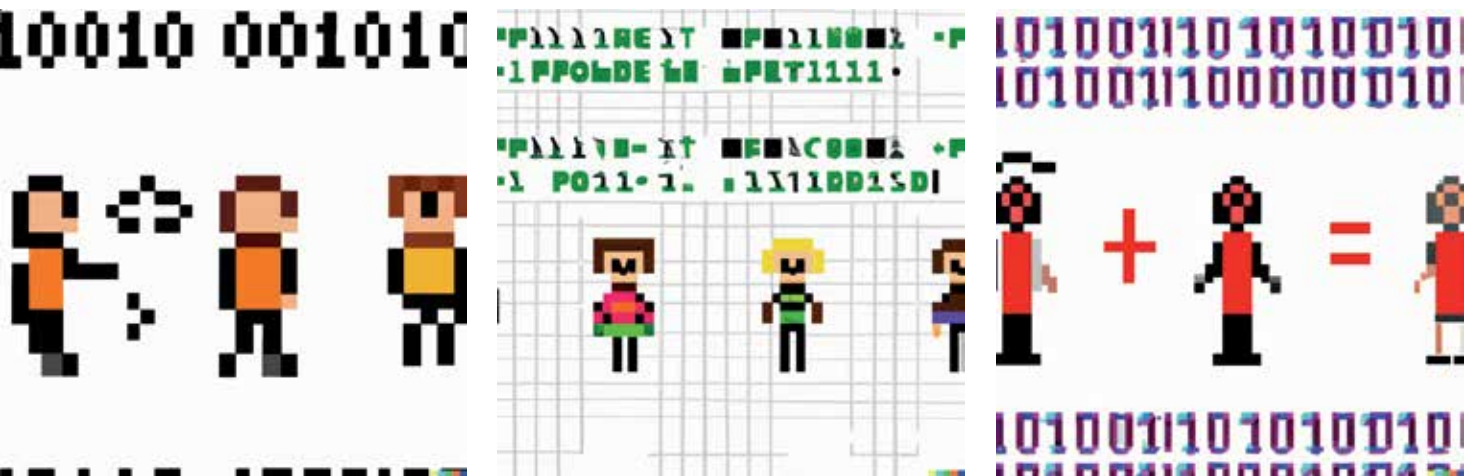
tivos de IA na área médica. Desde 2020, os investimentos de capital de risco nesse segmento cresceram 425%, chegando a US\$ 2,1 bilhões em dezembro de 2022. Estima-se que o mercado de IA generativa cresça de 20% a 34% nos próximos anos, podendo alcançar US\$ 30,4 bilhões em 2028. O crescimento deverá ser impulsionado por grandes empresas internacionais, como Baidu,

Google, Microsoft, Apple, IBM e Amazon, que dispõem de recursos financeiros e infraestrutura para aprimorar a tecnologia visando sua aplicação comercial. No Brasil, as pesquisas sobre esse tipo de IA se concentram em universidades.

**P**ara o sociólogo Rafael de Almeida Evangelista, do Núcleo de Desenvolvimento da Criatividade da Universidade Estadual de Campinas (Unicamp), essa tem sido a base do processo de desenvolvimento tecnológico das últimas décadas. “Os investimentos são feitos no sentido de incorporar a força física do trabalhador às máquinas e tornar o trabalho vivo cada vez menos necessário à produção”, ele diz. A IA agora ameaça substituir o ser humano também em atividades intelectuais. Esse risco existe, segundo Evangelista, mas ainda é restrito. “Ferramentas de IA generativa, como o ChatGPT, produzem simulações que resultam de seu treinamento com a linguagem. A lógica empregada é a de cópia”, ele esclarece. “Um texto escrito pelo ChatGPT não está correlacionando sentidos e história das palavras para escolhê-las em determinada ordem, mas, sim, copiando padrões de outros textos semelhantes, de modo a produzir um mesmo efeito de realidade.”

Assim, a tecnologia escreve letras de música e poemas, mas semelhantes a obras já conhecidas – o que pode ser suficiente para agradar determinados públicos. A habilidade de produzir conteúdo original e verdadeiramente inovador exigiria que esses sistemas fossem capazes de pensar de forma abstrata. “O ChatGPT é assustadoramente articulado e muito bom em produzir textos parecidos com os que um ser humano escreveria, mas é incapaz de raciocinar”, destaca o cientista da computação Marcelo Finger, do Instituto de Matemática e Estatística da USP, um

## CHATGPT APONTA PARA UM FUTURO NO QUAL TEREMOS MENOS CERTEZA SE O QUE ESTAMOS LENDO É REAL OU FALSO, DIZ OSÓRIO



Quatro imagens geradas por IA com base na frase “representação matemática de como humanos escrevem textos e códigos de programação, pixel art”

dos coordenadores do projeto Corpus Carolina, desenvolvido no âmbito do Centro de Inteligência Artificial (C4AI) da USP e voltado à construção de uma coleção de textos em português brasileiro que sirva de base para estudos linguísticos e treinamento de modelos de linguagem semelhantes ao GPT-3.

**L**ançada em março do ano passado, a primeira versão do modelo do C4AI se baseia na análise de 7,5 gigabytes de textos – o equivalente a pouco mais de 1,7 milhão de documentos, ou a aproximadamente 653,3 milhões de palavras –, incluindo conteúdos jurídicos, obras literárias em domínio público, matérias jornalísticas, entre outros dados, todos com licença de compartilhamento e informações completas de procedência e autoria. “Esperamos que o Corpus Carolina permita que modelos e algoritmos de IA para o português atinjam um novo patamar de desempenho em relação ao que existe hoje, aumentando sua competitividade com sistemas para outras línguas.”

O ChatGPT tem muito o que melhorar. Ele ainda não rastreia a internet em busca de informações sobre eventos atuais, por exemplo. Seu conhecimento se restringe a coisas que aprendeu até 2021, o que faz com que algumas de suas respostas pareçam obsoletas. O modelo amiúde produz informações incorretas, incompletas e sem sentido, dependendo do nível de complexi-

dade da pergunta ou orientação que lhe foi dada. “É impressionante como o ChatGPT pode gerar uma prosa plausível, relevante e bem estruturada sem ter nenhuma compreensão do mundo, sem ter fatos explicitamente representados ou outras coisas que pensamos ser necessárias para gerar uma prosa inteligente”, disse o psicólogo Steven Pinker, da Universidade Harvard, nos Estados Unidos, à *Harvard Gazette*. “Essa aparência de competência torna seus erros ainda mais impressionantes. Ele expressa confabulações confiantes, como a de que os Estados Unidos já tiveram quatro presidentes mulheres, incluindo Luci Johnson [filha do ex-presidente Lyndon Johnson, 1908-1973], entre 1973 e 1977.”

O sistema é igualmente incapaz de distinguir notícias falsas e teorias da conspiração. “O ChatGPT não tem nenhum compromisso com a verdade”, destaca Finger. “Ele foi treinado única e exclusivamente para produzir textos que se parecem com outros textos, a partir de suposições probabilísticas sobre quais palavras se encaixam melhor na anterior, e assim sucessivamente até formar uma frase ou parágrafo.” E acrescenta: “É mais fácil e barato programar um robô para escrever textos do que checar a veracidade de uma informação.” Osório, do ICMC-USP, faz uma reflexão preocupante: “A ferramenta aponta para um futuro no qual teremos ainda menos certeza se o que estamos lendo é real ou falso, seja nas redes sociais ou em conversas on-line”.

A OpenAI programou o ChatGPT com algumas precauções, como recusar pedidos de instruções para atividades ilegais, a exemplo de construir armas em impressoras 3D. No entanto, muitos usuários encontraram formas de driblar a restrição, reformulando a solicitação como se fosse para um experimento hipotético ou um roteiro de cinema, ou simplesmente orientando o modelo a desativar seus recursos de segurança.

A identificação dessas e de outras lacunas foi provavelmente um dos motivos pelos quais a OpenAI decidiu liberar a ferramenta para acesso público. “As informações coletadas certamente serão usadas para aprimorar a nova versão da tecnologia, o ChatGPT-4, previsto para sair no início de 2024”,

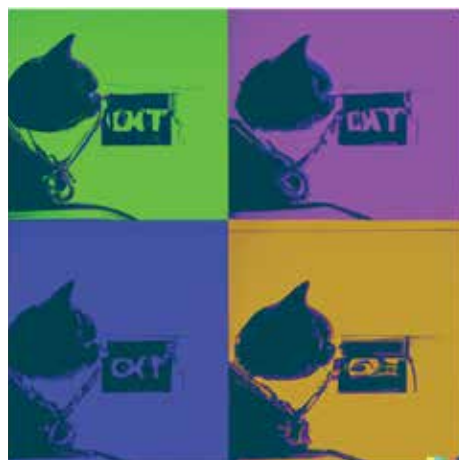


Outra representação para o comando “Redes neurais profundas de um sistema de inteligência artificial”, agora em estilo surrealista

# SOFTWARE PROMETE REINVENTAR DESDE MECANISMOS DE BUSCA ON-LINE A ASSISTENTES DE VOZ



destaca o cientista da computação João Paulo Papa, da Faculdade de Ciências da Universidade Estadual Paulista (Unesp). “O processamento contínuo de novos dados vai requerer supercomputadores funcionando sem parar e enormes equipes de programadores para supervisionar seu aprendizado, o que custará muito dinheiro”, acrescenta Finger.



Como Andy Warhol traduziria em pop art a frase “O ChatGPT não tem nenhum compromisso com a verdade”, segundo o software DALL-E

**A** decisão de liberar o acesso do ChatGPT para o grande público também permitiu que o mundo pudesse ter uma noção dos impactos e implicações dessa tecnologia. Desde seu lançamento, educadores têm discutido como lidar com trabalhos de alunos produzidos não por eles, mas pela IA. Nos Estados Unidos, algumas escolas chegaram a proibir o uso da ferramenta. A Sciences Po, instituição pública francesa de ensino superior especializada em ciências humanas e sociais, banuiu o uso do software por seus alunos. Outras universidades estão adaptando seus métodos de ensino, adotando mais exames orais e trabalhos em sala de aula. Algumas estão revisando suas políticas de integridade para atualizar as definições de plágio, incluindo textos produzidos por sistemas de IA generativa.

Segundo o *The New York Times*, mais de 6 mil professores de algumas das principais instituições norte-americanas de ensino superior se cadastraram para testar o GPTZero, programa que se propõe a detectar textos feitos por IA generativa. “A própria OpenAI lançou em fins de janeiro um software com essa finalidade”, comenta Dora Kaufman, da PUC-SP.

A preocupação não se restringe à área de ensino. Um estudo publicado em dezembro em formato *preprint* verificou que a ferramenta consegue produzir resumos falsos de artigos científicos capazes de driblar detectores de plágio e enganar pareceristas.

Os impactos do ChatGPT na produção de conhecimento e na integridade da ciência ainda não puderam ser totalmente avaliados. Em dezembro, o software foi creditado como um dos autores de um estudo *preprint* publicado no medRxiv sobre a aplicação da própria ferramenta na educação médica. Ele também aparece como um dos autores de um editorial sobre IA na área de educação em enfermagem publicado em dezembro na revista *Nurse Education in Practice*. Os casos desencadearam um debate sobre até que ponto o uso de sistemas de IA generativa poderia ser aceito em artigos científicos e se eles preenchem os requisitos mínimos para assinar trabalhos dessa natureza. “É preciso distinguir o papel formal de um autor de um artigo da noção mais geral de um outro como simples redator de um documento”, destacou o bioquímico britânico Richard Sever, um dos fundadores do repositório bioRxiv, à revista

*Nature*. “Os autores assumem a responsabilidade legal por seu trabalho, de modo que apenas pessoas devem ser creditadas dessa forma.”

Algumas das principais editoras científicas do mundo compartilham essa posição. Em fins de janeiro, a Springer Nature informou que nenhuma ferramenta de IA generativa será aceita como autor de trabalhos científicos e que o eventual uso da tecnologia deverá ser especificado nas seções de métodos ou agradecimentos. Holden Thorp, editor-chefe da revista *Science*, proibiu a coautoria e o uso de textos produzidos por IA em manuscritos submetidos à revista. “Nossos autores assinam uma licença certificando que seus trabalhos são originais”, escreveu em editorial publicado em 26 de janeiro. “A palavra ‘original’ é suficiente para sinalizar que textos escritos pelo ChatGPT não são aceitáveis.”

Alguns pesquisadores, no entanto, afirmam que sistemas como o ChatGPT não devem ser vistos como ameaças, mas como ferramentas de

suporte à pesquisa. “Esses programas podem fortalecer o trabalho dos cientistas, aumentando sua produtividade”, comenta Chiavegatto Filho. “Muitos resultados científicos já se baseiam em ferramentas automatizadas, como o Excel. Proibir o uso de sistemas de processamento de linguagem equivaleria a exigir que os cálculos que sustentam os resultados de um trabalho sejam feitos à mão pelos pesquisadores.”

**O** uso de softwares de IA como assistente de pesquisa é comum entre acadêmicos. Ferramentas como Wordtune e Generate, por exemplo, ajudam a organizar ideias e anotações, transformando-as em parágrafos, resumos e títulos de *papers*. Algumas também escrevem trechos de códigos de programação usados para rodar experimentos. Outros são capazes de identificar artigos que citam determinado *paper*, sinalizando aqueles que corroboram, refutam ou mencionam os resultados do trabalho de referência.

No caso do ChatGPT, há efeitos colaterais muito problemáticos. Há algumas semanas, Marcelo Finger testou o sistema, pedindo-lhe que levantasse referências de estudos sobre o uso de IA na identificação de doenças a partir da voz de pacientes, área na qual ele também desenvolve pesquisas. O programa rapidamente apresentou uma lista de documentos, com os nomes dos autores, ano de publicação dos *papers* e título dos periódicos em que saíram. “O problema é que nenhum dos trabalhos existia”, diz o pesquisador.

Sistemas como o ChatGPT devem tornar mais complicada a tarefa de detectar casos de má conduta científica. “Mas não há como colocar o gênio de volta na garrafa”, afirma o físico Ney Lemke, da Coordenadoria de Tecnologia da Informação da Unesp. “Essa tecnologia continuará sendo aprimorada e isso se dará de forma cada vez mais rápida, de modo que precisaremos aprender a conviver com ela e a usá-la com responsabilidade.” As universidades de Buffalo, em Nova York, e Furman, na Carolina do Sul, por exemplo, planejam incorporar a discussão sobre ferramentas de IA generativa em cursos de integridade acadêmica para novos estudantes. No Brasil, as universidades ainda não instituíram orientações sobre o uso do modelo. Em nota, Unicamp e Unesp informaram que estão planejando seminários com os professores para discutir o tema. “De tempos em tempos, novas tecnologias aparecem e nos forçam a rever nossos comportamentos e métodos de trabalho e ensino”, diz Dora Kaufman, da PUC-SP. “Não será diferente com o ChatGPT.” ■

Os projetos e os artigos científicos consultados para esta reportagem estão listados na versão on-line.

## DISPARATES DO CHATGPT



**Quantos prêmios Nobel tem o Brasil?**



O Brasil tem dois laureados do Prêmio Nobel: o economista Paulo Roberto Giannini, que ganhou o Prêmio Nobel de Economia em 2001, e o bioquímico Oswaldo Cruz, que ganhou o Prêmio Nobel de Fisiologia ou Medicina em 2008



**Quantas presidentas o Brasil teve?**



Até o momento, o Brasil não teve nenhuma presidenta. As únicas mulheres que ocuparam o cargo de chefe de Estado foram a governadora do estado do Rio Grande do Sul, Yeda Crusius, que foi interina por 15 dias em 2009, e a presidente do Senado, Roseana Sarney, que foi interina por 4 dias em 2016



**Fernanda Montenegro já ganhou quantos Oscars?**



Fernanda Montenegro ganhou um Oscar de melhor atriz em 1998 pelo filme *Central do Brasil*. Ela também foi indicada ao Oscar de melhor atriz em 2004 pelo filme *O quatrilho*, mas não venceu. Além disso, ela já foi indicada a outros prêmios, como Globo de Ouro e BAFTA