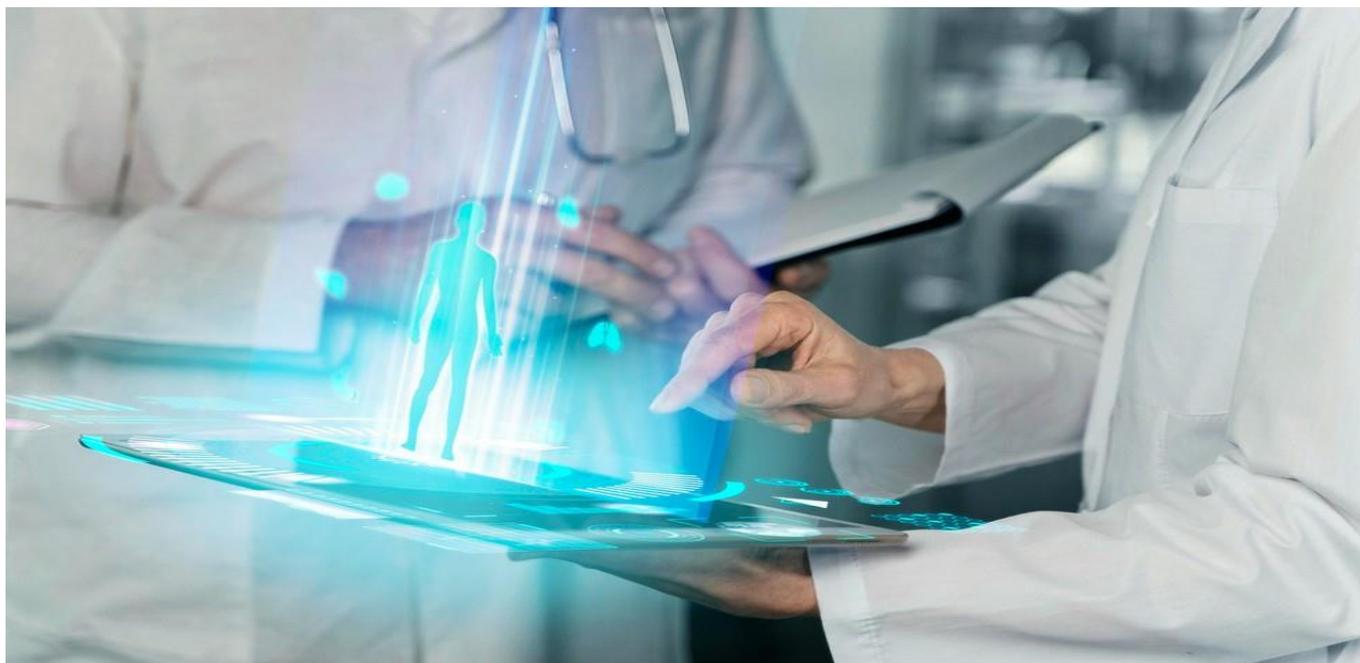


Inteligência artificial já pode ter consciência? O que dizem os cientistas

Avanço dos grandes modelos de linguagem têm levado especialistas em tecnologia a se perguntar se já estamos vivendo na era em que o que antes era ficção científica se tornou a realidade.

Por BBC

26/05/2025 06h02 Atualizado há um mês



Avanço dos grandes modelos de linguagem têm levado especialistas em tecnologia a se perguntar se já estamos vivendo na era em que o que antes era ficção científica se tornou a realidade. — Foto: Freepik
 Entro na cabine com alguma apreensão. A qualquer instante, vou ser exposto a uma luz estroboscópica enquanto uma música toca no fundo, tudo parte de um projeto de pesquisa que tenta entender o que nos torna verdadeiramente humanos.

É uma experiência que lembra o teste do filme de ficção científica *Blade Runner*, criado para distinguir humanos de seres criados artificialmente que se passam por humanos.

Será que eu poderia ser um robô do futuro e não saber? Passaria no teste?

Os pesquisadores me garantem que não é disso que se trata o experimento. O dispositivo que eles chamam de "máquina dos sonhos" foi projetado para estudar como o cérebro humano gera nossas experiências conscientes do mundo.

Quando o estroboscópio começa, mesmo com os olhos fechados, vejo padrões geométricos bidimensionais em espiral.

É como pular em um caleidoscópio, com triângulos, pentágonos e octógonos em constante mudança. As cores são vivas, intensas e em constante mudança: tons de rosa, magenta e turquesa, brilhando como luzes de neon.

A "máquina dos sonhos" traz a atividade interna do cérebro à tona com luzes piscantes com o objetivo de explorar como nossos processos de pensamento funcionam.

As imagens que vejo são exclusivas do meu mundo interior e únicas para mim, de acordo com os pesquisadores, que acreditam que esses padrões podem lançar luz sobre a própria consciência.

Eles me ouvem sussurrar: "É lindo, absolutamente lindo. É como voar pela minha própria mente!"

A "máquina dos sonhos", no Centro de Ciência da Consciência da Universidade de Sussex, é apenas um dos muitos novos projetos de pesquisa ao redor do mundo que investigam a consciência humana: a parte de nossas mentes que nos permite ter autoconsciência, pensar, sentir e tomar decisões independentes sobre o mundo.

Ao aprender sobre a natureza da consciência, os pesquisadores esperam entender melhor o que está acontecendo dentro dos cérebros de silício da inteligência artificial.

Alguns acreditam que os sistemas de IA em breve se tornarão conscientes de forma independente, se é que ainda não o são.

Mas o que realmente é consciência e quão perto a IA está de obtê-la? E a crença de que a IA pode ser consciente pode mudar fundamentalmente os humanos nas próximas décadas?

Da ficção científica à realidade

A ideia de máquinas com mentes próprias tem sido explorada há muito tempo na ficção científica. As preocupações com a IA remontam a quase cem anos, ao filme *Metrópolis*, no qual um robô se passa por uma mulher real.

O medo de que as máquinas se tornassem conscientes e representassem uma ameaça aos humanos foi explorado no filme *2001: Uma Odisseia no Espaço*, de 1968, quando o computador HAL 9000 tentou matar astronautas a bordo de sua nave espacial.

E no último filme *Missão Impossível*, que acaba de ser lançado, o mundo é ameaçado por uma poderosa IA desonesta, descrita por um personagem como um "parasita digital autoconsciente, autodidata e devorador de verdades".

No mundo real, contudo, muito recentemente houve uma rápida mudança no pensamento sobre a consciência das máquinas — especialistas começaram a expressar preocupação de que isso não seja mais coisa de ficção científica.

A mudança repentina foi motivada pelo sucesso dos chamados grandes modelos de linguagem (LLMs, na sigla em inglês), que podem ser acessados por meio de aplicativos como Gemini e Chat GPT.

A capacidade da última geração dos LLMs de ter conversas plausíveis e fluidas surpreendeu até mesmo seus criadores e alguns dos principais especialistas da área.

Há uma visão crescente entre alguns pensadores de que, à medida que a IA se torna ainda mais inteligente, as luzes se acenderão repentinamente dentro das máquinas e elas se tornarão conscientes.

Outros, como o professor Anil Seth, que lidera a equipe da Universidade de Sussex, discordam, descrevendo a visão como "cegamente otimista e motivada pelo excepcionalismo humano".

Associamos consciência à inteligência e à linguagem porque elas andam juntas nos humanos. Mas só porque andam juntas em nós, não significa que andem juntas em geral, por exemplo, nos animais.

Então, o que é realmente consciência?

A resposta curta é que ninguém sabe. Isso fica claro pelos argumentos bem-humorados, porém robustos, da própria equipe do professor Seth, composta por jovens especialistas em IA, especialistas em computação, neurocientistas e filósofos, que estão tentando responder a uma das maiores questões da ciência e da filosofia.

Embora existam muitas visões diferentes no centro de pesquisa da consciência, os cientistas estão unidos em seu método: dividir esse grande problema em muitos outros menores em uma série de projetos de pesquisa, que inclui a "máquina dos sonhos".

Assim como a busca pela "centelha de vida" que poderia dar vida a objetos inanimados foi abandonada no século 19 em favor da identificação de como partes individuais dos sistemas vivos funcionavam, a equipe de Sussex agora está adotando a mesma abordagem para a consciência.

Eles esperam identificar padrões de atividade cerebral que expliquem várias propriedades de experiências conscientes, como mudanças em sinais elétricos ou fluxo sanguíneo para diferentes regiões.

O objetivo é ir além da busca por meras correlações cerebrais entre atividade e consciência e tentar encontrar explicações para seus componentes individuais.

O professor Seth, autor do livro sobre consciência *Being You* ("sendo você", em tradução literal), teme que possamos estar nos precipitando em uma sociedade que está sendo rapidamente remodelada pelo ritmo acelerado das mudanças tecnológicas sem conhecimento suficiente sobre a ciência ou reflexão sobre as consequências.

"Entendemos isso como se o futuro já estivesse escrito; que há uma marcha inevitável para uma substituição sobre-humana", diz ele.

"Não tivemos essas conversas o suficiente com o surgimento das mídias sociais, para nosso prejuízo coletivo. Mas com a IA, ainda não é tarde demais. Podemos decidir o que queremos."

Já existe consciência na inteligência artificial?

Alguns no setor de tecnologia acreditam que a IA em nossos computadores e telefones pode já estar consciente e que, por isso, devemos tratá-la como tal.

O Google suspendeu o engenheiro de software Blake Lemoine em 2022 após ele argumentar que chatbots de inteligência artificial podiam sentir e potencialmente sofrer.

Em novembro de 2024, um responsável pelo bem-estar da IA da Anthropic, Kyle Fish, foi coautor de um relatório sugerindo que a consciência da IA era uma possibilidade realista em um futuro próximo. Ele disse recentemente ao *The New York Times* que também acreditava que havia uma pequena chance (15%) de que os chatbots já estivessem conscientes.

Um dos motivos pelos quais ele acredita que isso seja possível é que ninguém, nem mesmo as pessoas que desenvolveram esses sistemas, sabem exatamente como eles funcionam.

Isso é preocupante, diz o professor Murray Shanahan, cientista do Google DeepMind e professor emérito de IA no Imperial College, em Londres.

"Na verdade, não entendemos muito bem a maneira como os LLMs (sigla para "large language models", ou grandes modelos de linguagem) trabalham internamente, e isso é motivo de preocupação", declarou à BBC.

De acordo com o professor Shanahan, é importante que as empresas de tecnologia tenham uma compreensão adequada dos sistemas que estão construindo — e os pesquisadores estão analisando isso com urgência.

"Estamos em uma posição estranha ao construir essas coisas extremamente complexas, onde não temos uma boa teoria de como exatamente elas alcançam os feitos extraordinários que estão alcançando", diz ele.

"Portanto, entender melhor como eles funcionam nos permitirá orientá-los na direção que queremos e garantir que estejam seguros."

'O próximo passo na evolução da humanidade'

A visão predominante no setor de tecnologia é que os LLMs não têm consciência da maneira como vivenciamos o mundo atualmente, e provavelmente não têm consciência alguma. Mas isso é algo que o casal de professores Lenore e Manuel Blum, ambos professores eméritos da Universidade Carnegie Mellon em Pittsburgh, no Estado americano da Pensilvânia, acredita que mudará, possivelmente em breve.

De acordo com os Blum, isso pode acontecer à medida que a IA e os LLMs tiverem mais entradas sensoriais ao vivo do mundo real, como visão e tato, ao conectar câmeras e sensores táteis (relacionados ao toque) aos sistemas de IA.

Eles estão desenvolvendo um modelo de computador que constrói sua própria linguagem interna, chamada "Brainish", para permitir que esses dados sensoriais adicionais sejam processados, tentando replicar os processos que ocorrem no cérebro.

"Acreditamos que o 'Brainish' pode resolver o problema da consciência como a conhecemos", diz Lenore à BBC. "A consciência da IA é inevitável."

Manuel intervém com entusiasmo, dizendo que os novos sistemas que ele acredita firmemente que surgirão serão o "próximo estágio na evolução da humanidade".

Robôs conscientes, ele acredita, "são nossa progênie". "No futuro, máquinas como essas serão entidades que estarão na Terra e talvez em outros planetas quando não estivermos mais aqui."

David Chalmers, professor de Filosofia e Ciência Neural na Universidade de Nova York, definiu a distinção entre consciência real e aparente em uma conferência em Tucson, Arizona, em 1994. Ele expôs o "difícil problema" de descobrir como e por que qualquer uma das operações complexas do cérebro dá origem à experiência consciente, como nossa resposta emocional quando ouvimos um rouxinol cantar.

Ele diz que está aberto à possibilidade de que o difícil problema seja resolvido.

"O resultado ideal seria aquele em que a humanidade compartilhasse dessa nova bonança de inteligência", afirma à BBC. "Talvez nossos cérebros sejam aprimorados por sistemas de IA."

Sobre as implicações disso para a ficção científica, ele observa ironicamente: "Na minha profissão, há uma linha tênue entre ficção científica e filosofia."

'Computadores feitos de carne'

O professor Seth, no entanto, está explorando a ideia de que a verdadeira consciência só pode ser alcançada por sistemas vivos.

"Pode-se argumentar fortemente que não é a computação que é suficiente para a consciência, mas estar vivo", diz ele.

"Nos cérebros, diferentemente dos computadores, é difícil separar o que eles fazem do que eles são."

Sem essa separação, ele argumenta, é difícil acreditar que os cérebros "são simplesmente computadores feitos de carne".

E se a intuição do professor Seth sobre a importância da vida estiver no caminho certo, a tecnologia mais provável de ser desenvolvida nesse sentido não será feita de silício e executada em código de programação, mas consistirá nas pequenas coleções de células nervosas do tamanho de grãos de lentilha que estão atualmente sendo cultivadas em laboratórios.

Apelidados de "minicérebros" pela imprensa, eles são chamados de "organoides cerebrais" pela comunidade científica, que os utiliza para pesquisar como o cérebro funciona e para testes relacionados a medicamentos.

Uma empresa australiana chamada Cortical Labs desenvolveu um sistema de células nervosas em uma placa de vidro capaz de jogar o videogame esportivo Pong, de 1972.

Embora esteja muito longe de ser um sistema consciente, o chamado "cérebro em uma placa" é assustador, pois move uma raquete para cima e para baixo na tela para rebater uma bola pixelada.

Alguns especialistas acreditam que, se a consciência surgir, é mais provável que venha de versões maiores e mais avançadas desses sistemas de tecidos vivos.

O Cortical Labs monitora sua atividade elétrica em busca de quaisquer sinais que possam ser evidência de algo como o surgimento da consciência.

O diretor científico e operacional da empresa, Brett Kagan, está ciente de que qualquer inteligência emergente incontrolável pode ter prioridades que "não estão alinhadas com as nossas".

Nesse caso, ele diz, um pouco em tom de brincadeira, que possíveis vilões organoídes seriam mais fáceis de derrotar porque "sempre haverá água sanitária" para despejar sobre seus neurônios frágeis.

Voltando a um tom mais solene, ele afirma que a pequena, mas significativa ameaça da consciência artificial é algo em que gostaria que os grandes nomes da área se concentrassem mais como parte de tentativas sérias de avançar nossa compreensão científica.

No entanto, acrescenta, "infelizmente, não vemos nenhum esforço sério nesse espaço".

A ilusão de consciência

O problema mais imediato, porém, pode ser como a ilusão de que as máquinas são conscientes nos afeta. Em apenas alguns anos, poderemos estar vivendo em um mundo povoado por robôs humanoides e deepfakes que parecem conscientes, pondera o professor Seth. Ele se preocupa que não seremos capazes de resistir à crença de que a IA tem sentimentos e empatia, o que pode levar a novos perigos.

"Isso significa que confiamos mais nessas coisas, compartilhamos mais dados com elas e estamos mais abertos à persuasão."

Mas o maior risco da ilusão da consciência é uma "corrosão moral", diz ele.

"Isso distorcerá nossas prioridades morais, fazendo com que dediquemos mais recursos a cuidar desses sistemas em detrimento das coisas reais em nossas vidas" — o que significa que podemos ter compaixão por robôs, mas nos importar menos com outros humanos.

E isso poderia nos mudar fundamentalmente, afirma o professor Shanahan.

"Cada vez mais, os relacionamentos humanos serão replicados em relacionamentos de IA; eles serão usados como professores, amigos, adversários em jogos de computador e até mesmo parceiros românticos. Se isso é bom ou ruim, não sei, mas vai acontecer, e não seremos capazes de impedir."

FONTE: <https://www.bbc.com/portuguese/articles/cvgn9816jmeo>