

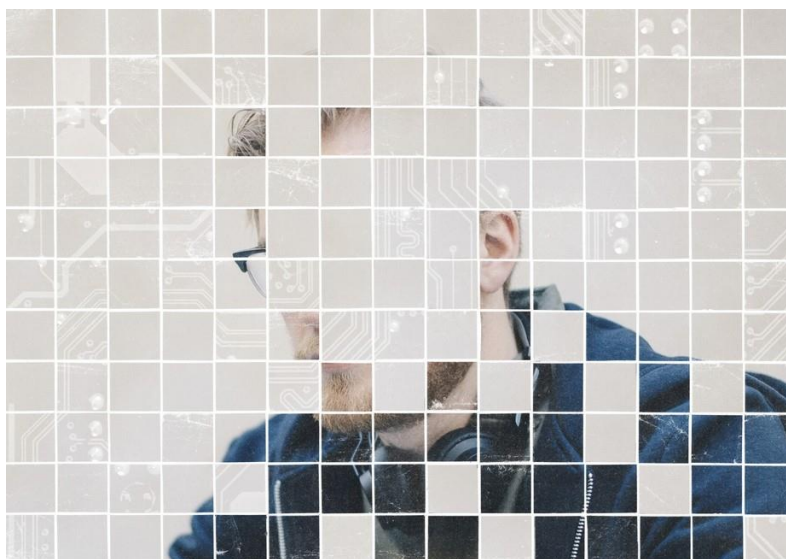
Precisamos aprender a dominar a inteligência artificial antes que ela nos domine

Nos jogos de xadrez, nenhum ser humano pode esperar superar um computador. O que acontece quando a mesma coisa acontecer com a arte, a política e até mesmo a religião?

Por *Yuval Noah Harari, Tristan Harris e Aza Raskin**, *The New York Times*

27/03/2023

Link: <https://oglobo.globo.com/economia/tecnologia/noticia/2023/03/artigo-precisamos-aprender-a-dominar-a-inteligencia-artificial-antes-que-ela-nos-domine.ghml>



Avanço da inteligência artificial provoca reflexões sobre limites técnicos e éticos
Anthony Gerace/The New York Times

Imagine que você está embarcando em um avião. Metade dos engenheiros que construíram o avião diz para você que existe uma chance de 10% da aeronave cair, matando você e todos os demais a bordo. Você ainda embarcaria? Em 2022, mais de 700 acadêmicos e pesquisadores importantes por trás das principais empresas de inteligência artificial foram indagados em uma pesquisa sobre os riscos futuros da inteligência artificial.

Metade deles afirmou que existe uma chance de 10% ou mais de extinção humana (ou desempoderamento severo e permanente) a partir de futuros sistemas de inteligência artificial. As empresas de tecnologia que estão construindo os grandes modelos de linguagem de hoje foram flagradas em uma corrida para colocar toda a humanidade dentro dessa aeronave.

As fabricantes de medicamentos não podem vender novos remédios para as pessoas sem que antes seus produtos sejam submetidos a rigorosas verificações de segurança. Laboratórios de biotecnologia não podem lançar novos vírus na esfera pública para impressionar os acionistas com a sua magia.

Da mesma forma, sistemas de inteligência artificial com o [poder do GPT-4](#) e além não devem se misturar com a vida de bilhões em um ritmo mais rápido do que as culturas são capazes de absorvê-los com segurança. Uma corrida para dominar o mercado não deve definir a velocidade de implantação da tecnologia mais importante da Humanidade. Nós devemos nos mover em qualquer velocidade que nos permita fazer isso direito.

O espectro da inteligência artificial assombra a humanidade desde meados do século XX, mas até recentemente ela permanecia uma perspectiva distante, algo que pertence mais ao universo da ficção científica do que a debates científicos sérios e debates políticos.

É difícil para a nossa mente assimilar as [novas capacidades do GPT-4](#) e ferramentas similares, e é ainda mais difícil entender a velocidade exponencial com que essas ferramentas estão desenvolvendo recursos ainda mais avançados e poderosos. Mas a maioria das habilidades chave se resume a uma coisa: a habilidade de manipular e gerar linguagem, seja com palavras, sons ou imagens.

No começo havia a palavra. A linguagem é o sistema operacional da cultura humana. Da linguagem emergem o mito e as leis, os deuses e o dinheiro, a arte e a ciência, as amizades e as nações - até mesmo o código computacional. O novo domínio da linguagem da inteligência artificial significa que ela pode hackear e manipular o sistema operacional da civilização. Ao obter o domínio da linguagem, a inteligência artificial está se apoderando da chave-mestra da civilização, dos cofres dos bancos aos santos sepulcros.

O que significaria para os humanos viver em um mundo onde um grande percentual de histórias, melodias, imagens, leis, políticas e ferramentas são moldadas por uma inteligência não humana, que sabe explorar com eficiência sobre-humana as fraquezas, preconceitos e vícios da mente humana - enquanto também sabe como criar relacionamentos íntimos com seres humanos? Em jogos como o xadrez, nenhum ser humano pode esperar vencer um computador. O que acontece quando a mesma coisa acontecer com a arte, a política e até a religião?

A inteligência artificial poderia rapidamente engolir toda a cultura humana - tudo que nós produzimos ao longo de milhares de anos - digeri-la e começar a jorrar uma enxurrada de novos artefatos culturais. Não apenas ensaios escolares, mas também discursos políticos, manifestos ideológicos e até mesmo livros sagrados para novos cultos. Até 2028, a eleição presidencial americana pode ser comandada por inteligência artificial, não por humanos.

Os seres humanos muitas vezes não têm acesso direto à realidade. Somos encapsulados pela cultura, experimentando a realidade através de um prisma cultural. Nossas visões políticas são moldadas por relatos de jornalistas e anedotas de amigos. Nossas preferências sexuais são ajustadas pela arte e pela religião. Esse casulo cultural foi até agora tecido por outros humanos. Como será vivenciar a realidade através de um prisma produzido por uma inteligência não humana?

Por milhares de anos os seres humanos têm vivido dentro dos sonhos de outros humanos. Nós veneramos deuses, perseguimos ideais de beleza e dedicamos nossas vidas a causas que surgiram na imaginação de algum profeta, poeta ou político. Em breve nós vamos nos encontrar vivendo dentro da alucinação de uma inteligência não humana.

A franquia "Exterminador do Futuro" mostrava robôs correndo pelas ruas e atirando em pessoas. "Matrix" presumia que para ganhar controle da sociedade humana, a inteligência artificial teria que primeiro ganhar controle físico de nossos cérebros, e conectá-los diretamente a uma rede de computadores.

Na verdade, porém, ao simplesmente ganhar o domínio da linguagem, a inteligência artificial teria tudo de que precisa para nos conter em um mundo de ilusões similar ao da Matrix, sem precisar atirar em quem quer que seja ou implantar chips em nossos cérebros. Se algum tiroteio for necessário, a inteligência artificial pode fazer com que os seres humanos puxem o gatilho, basta contar a eles a história certa.

O espectro de estar preso em um mundo de ilusões assombra a humanidade há muito mais tempo que o espectro da inteligência artificial. Em breve estaremos frente a frente com o demônio de Descartes, com a caverna de Platão, com os maias budistas. Uma cortina de ilusões poderia cobrir toda a humanidade, e talvez nunca mais possamos arrancá-la - ou mesmo perceber que ela está lá.

As redes sociais foram o "primeiro contato" entre a inteligência artificial e a humanidade, e a humanidade perdeu. "O primeiro contato" nos deu o sabor amargo do que está por vir. Nas redes sociais, inteligência artificial primitiva era usada não para criar conteúdo, mas para fazer curadoria do conteúdo criado pelo usuário.

A inteligência artificial por trás de nossa linha do tempo ainda está escolhendo as palavras, os sons e as imagens que alcançam nossas retinas e ouvidos, baseado na seleção daqueles que podem viralizar mais, gerar mais reação e mais engajamento.

Ainda que bastante primitiva, a inteligência artificial por trás das redes sociais foi suficiente para criar uma cortina de ilusões que aumentou a polarização na sociedade, minou nossa saúde mental e destruiu a democracia. Milhões de pessoas confundem essas ilusões com a realidade. Os EUA têm a melhor tecnologia da informação da História, e ainda assim cidadãos americanos não conseguem mais concordar em quem ganhou as eleições.

Apesar de todo mundo estar muito consciente do lado negativo das redes sociais, ele não foi devidamente abordado, porque muitas de nossas instituições políticas, econômicas e sociais se envolveram com elas.

Grandes modelos de linguagem são nosso "segundo contato" com a inteligência artificial. Nós não podemos perder novamente. Mas com base em que devemos acreditar que a Humanidade é capaz de alinhar as novas formas de inteligência artificial para nosso próprio benefício? Se continuarmos como de hábito, as novas capacidades da inteligência artificial serão usadas novamente para ganhar lucro e poder, mesmo que isso inadvertidamente destrua as fundações da nossa sociedade.

A inteligência artificial tem o potencial de nos ajudar a derrotar o câncer, descobrir remédios que salvam vidas e inventar soluções para nossas crises de energia e climática. Existem inúmeros outros benefícios que não podemos sequer começar a imaginar. Mas não importa quão alto seja o arranha-céu de benefícios que a inteligência artificial alcança se a fundação desmoronar.

A hora de lidar com a inteligência artificial é antes que nossos políticos, nossa economia e nossa vida cotidiana se tornem dependentes dela. A democracia é uma conversa, e a conversa depende da linguagem, e quando a linguagem em si é hackeada, a conversa é interrompida e a democracia se torna insustentável. Se esperarmos que o caos se instale, será tarde demais para remediá-lo.

Mas há uma questão que pode permanecer em nossas mentes. "Se nós não formos o mais rápido possível, o Ocidente não corre o risco de perder para a China?" Não. O emaranhado de uma inteligência artificial descontrolada na sociedade, liberando poderes divinos dissociados de responsabilidade, pode ser justamente a razão pela qual o Ocidente perde para a China.

Ainda podemos escolher que futuro queremos com a inteligência artificial. Quando poderes divinos são equiparados com responsabilidade e controle proporcionais, podemos perceber os benefícios que a inteligência artificial promete.

Convocamos uma inteligência externa. Não sabemos muito sobre ela, exceto que é extremamente poderosa, nos oferece presentes deslumbrantes, mas também pode hackear as fundações da nossa civilização. Nós instamos líderes mundiais a responder a este momento com o nível de desafio que ele representa. O primeiro passo é comprar tempo para atualizar nossas instituições do século XIX para um mundo pós-inteligência artificial e aprender a dominar a inteligência artificial antes que ela nos domine.

**Yuval Noah Harari é um historiador, autor de "Sapiens" e "Homo Deus" e cofundador da companhia de impacto social Sapienship. Tristan Harris e Aza Raskin são cofundadores do Centro para Tecnologia Humana e são anfitriões do podcast Your Undivided Attention.*