

Perigos da 'inteligência' artificial¹

Valdemar W. Setzer

Depto. de Ciência da Computação, IME-USP

www.ime.usp.br/~vwsetzer – Esta versão: 10/5/25

Os leitores devem ter estranhado o emprego de *inteligência* entre aspas no título. Ocorre que a ciência não sabe o que é inteligência e como ela se processa: assim, a denominação 'inteligência' artificial (IA) é errada. Um nome mais apropriado seria 'simulação digital de processos cognitivos', que obviamente jamais será adotada. A abreviatura IA será aqui empregada por ser popular. Esse erro no nome já representa um perigo, pois dá a impressão de que algo artificial pode ser como a nossa inteligência e, com isso, dar a impressão de sermos máquinas, o que pode trazer terríveis consequências (por exemplo, máquinas não têm empatia nem compaixão, como será exposto logo adiante). Por outro lado, eleva os computadores a um nível indevido, achando-se que haverá máquinas com todas as capacidades humanas, já que a inteligência é uma das mais elevadas.

Duas dessas capacidades jamais estarão em máquinas: nossas sensações e sentimentos, pois são individuais e subjetivos, ao passo que computadores são objetivos e universais – com capacidade suficiente, qualquer um pode simular exatamente ('emular') qualquer outro. Isso significa que se pode dar como entrada a um computador A o código interno (a assim chamada 'linguagem de máquina', que não é usada na programação) de um programa que seria executado por um computador B. Um interpretador desse código de B é executado em A, com os dados de entrada que seriam dados ao B, chegando exatamente aos mesmos dados de saída.

Se uma pessoa come um caqui e tem a sensação do gosto dele, nenhuma outra pessoa pode ter a mesma sensação; se a pessoa gosta ou não gosta do caqui, esse sentimento também não pode ser sentido por outra pessoa, daí a individualidade e subjetividade do ser humano. Em *A Filosofia da Liberdade*, no capítulo VI "A individualidade humana", Rudolf Steiner salienta um aspecto fundamental que sensações e sentimentos dão ao ser humano: "Nosso pensar nos une ao mundo; nosso sentir nos conduz a nós mesmos, só assim fazendo de nós um indivíduo. Se nós fôssemos apenas seres pensantes e perceptivos, nossa vida inteira teria que transcorrer numa indiferença sem distinções." [Steiner 2022]. Os aspectos individuais e subjetivos dos seres humanos já mostram que máquinas jamais terão todas as capacidades humanas. Sobre sentimentos e sensações, ver meu artigo "Conceitos, ideias, sensações, sentimentos e a cognição" [Setzer 2024], onde faço inicialmente uma clara distinção entre conceitos e ideias, e mostro que ideias não podem estar armazenadas no cérebro ou em máquinas.

Ultimamente, com o advento da IA generativa, os resultados da IA tornaram-se muito bons em termos linguísticos e imagéticos, tornando-se convincentes. Assim, pessoas não especializadas podem ser enganadas com resultados falsos ou viciados, um grande perigo de formar opiniões erradas ou tendenciosas. Qualquer usuário da internet já tem um perfil de seus dados armazenado em vários sistemas: duas pessoas fazendo a mesma pergunta para uma máquina de busca ou para um dado sistema de IA generativa provavelmente terão respostas diferentes, baseadas em seus perfis. Com isso, resultados dessa IA podem ser personalizados, tornando-se ainda mais convincentes. Com isso, pessoas desavisadas podem achar que os computadores terão todas as capacidades intelectuais de um ser humano. Nesse sentido, os computadores são um grande instrumento para propagar o materialismo e o cientificismo, isto é, a crença na ciência e nos cientistas. Ver sobre isso a apresentação de minha palestra "O computador com instrumento do cientificismo" [Setzer 2025].

Um caso particular são os tradutores automáticos. Qualquer pessoa que os usou sabe que seu resultado deve ser revisto por alguém que conheça bem a língua original e a da tradução. Eles são úteis como primeira aproximação, mas não conterão o estilo próprio do tradutor humano.

É sintomático que os computadores se tornaram a maior arma de enganar pessoas e provocar fraudes. Como as fraudes pela internet crescem exponencialmente, e dados úteis crescem linearmente, já houve previsões de que, como tudo o que cresce exponencialmente acaba explodindo, a internet ainda se tornará inviável.

Um perigo muito grande é o fato de a IA generativa usar dados que foram fornecidos durante a fase de 'treinamento' – novamente um nome errado, pois somente seres humanos e animais podem ser treinados. Nessa fase é dada ao sistema uma quantidade enorme de dados de entrada e os correspondentes esperados na saída. Com isso, usando enormes capacidades computacionais e técnicas matemáticas complexas, são calculados muitos parâmetros (hoje em dia, milhões deles; há uma notícia de que o ChatGPT4 usa 1,5 trilhão de parâmetros).

Por exemplo, pode-se 'treinar' (mais uma nomenclatura antropomórfica indevida; melhor seria dizer 'parametrizar') um desses sistemas para distinguir fotos de cães e de gatos. Fornece-se como entrada milhares de fotos, cada uma com o resultado esperado, isto é, 'cão' ou 'gato'. Com isso, num processo de otimização (diminuição do erro), o sistema calcula milhares, milhões de parâmetros que processam uma nova foto de entrada, dando como resultado 'cão' ou 'gato', com uma alta probabilidade de acerto. Dando-se uma foto de raposa ou lobo, certamente o resultado será 'cão', mostrando como o sistema é parcial e viciado. 'Treinando-se' o sistema com fotos de temas diferentes, ele poderá dar respostas

a outros tópicos, cada um com um novo algoritmo, ou melhor, parâmetros específicos para cada tópico. Os grandes sistemas generativos de linguagem analisam enormes quantidades de textos, respondendo perguntas baseando-se na análise de textos com palavras ou frases semelhantes aos dados das perguntas. Trata-se de um sistema estatístico, que calcula a resposta com a maior probabilidade de ser bem adaptada aos textos existentes na máquina. Na verdade, as palavras não ficam armazenadas nas máquinas, mas símbolos representativos delas, chamados de *Tokens*. Um *token* pode ser, por exemplo, um prefixo ou sufixo, usando os mesmos para várias palavras, diminuindo assim os dados armazenados. As notícias são de que o ChatGPT4 usa 32.000 *tokens* de contexto para compor a frase mais adequada.

Os sistemas generativos representam uma revolução na computação: na programação tradicional de computadores, um programador sabia exatamente, matematicamente, o que cada comando de seu programa iria forçar a máquina a executar. Assim, o programa era teoricamente conhecido. No entanto, se o programa contivesse erros, os resultados poderiam ser imprevisíveis, o que era detectado por meio de testes. É interessante notar que um teste com certos dados de entrada pode revelar que, para esses dados, o programa funciona, fornecendo os resultados esperados. No entanto, isso não garante que o programa vai funcionar para outros dados de entrada. Em lugar de usar testes, poder-se-ia provar matematicamente que um programa está correto, como se fosse um teorema. No entanto, isso é praticamente impossível para programas de porte razoável. Por isso tantos programas recebem atualizações: durante o uso, erros foram descobertos que não tinham sido detectados inicialmente.

A IA generativa é um sistema de gerar programas, modificando seus parâmetros para processar dados com diferentes propósitos. O grande problema é que, como a quantidade de parâmetros calculados em cada caso é imensa, é impossível que eles sejam examinados e se compreenda como o programa funciona. Assim, os computadores adquiriram certa independência: eles mesmos geram seus programas (ou melhor, modificam seus programas, com base nos dados fornecidos), cujo funcionamento ninguém conhece, ninguém compreende. Temos aí o maior perigo representado pela IA generativa.

Isso não é muito grave quando computadores controlam máquinas, como pilotos automáticos em aviões, ou automóveis autônomos pois, infelizmente, a grande maioria dos acidentes é provocada por falhas humanas. Segundo o Observatório Nacional de Segurança Viária, 90% dos acidentes de trânsito são devidos a falhas humanas. Se sistemas autônomos diminuíssem muito esses desastres, deverão ser usados. É interessante notar que, atualmente, sistemas de carros autônomos não preveem todos os possíveis casos do que pode ocorrer à volta do veículo – daí alguns desastres que têm acontecido.

A situação torna-se muito grave quando os resultados da IA são usados para substituir decisões humanas, que afetam outras pessoas ou a sociedade. Nesse caso, deveria haver leis que proibissem o uso de resultados da IA sem que seres humanos examinem esses resultados e aprovem os resultados. Em outras palavras, um resultado que influencia diretamente pessoas deveria ser usado exclusivamente como sugestão. Um exemplo disso é o diagnóstico médico. Aliás, aqui há um outro grande perigo: os médicos confiarem nesses diagnósticos e perderem a capacidade de fazer os seus. Computadores só fazem escolhas lógico-matemáticas. Somente seres humanos tomam decisões, que podem, por exemplo, ser baseadas em empatia e compaixão, algo que jamais as máquinas poderão exercer, como exposto acima.

Outro grande perigo é o congelamento da evolução cultural humana. Os dados usados pela IA para chegar aos seus resultados são sempre dados do passado. Obviamente computadores não têm a intuição humana, que pode e deve trazer inovações para o futuro, que não são simplesmente combinação de coisas passadas.

Finalmente, é importante saber que computadores somente processam dados quantificados. Internamente, os dados são armazenados em algum sistema numérico; é usado o sistema binário em todos os computadores por pura conveniência tecnológica, pois nesse sistema numérico os circuitos eletrônicos ficam mais simples, rápidos e confiáveis. Usando um sistema binário, é possível representar um número em qualquer outro sistema. Por exemplo, o binário 10101 é equivalente ao decimal 21. O que não pode ser quantificado, não pode ser fornecido a um computador. Note-se que as letras do alfabeto já são naturalmente quantificadas, havendo 26 delas, podendo-se associar a cada uma um código numérico, isto é, binário. Por isso computadores podem processar textos. Uma foto pode ser quantificada, associando-se números à posição de cada elemento da foto ('píxel') e à sua cor dada, por sinal, pelas intensidades do vermelho, do verde e do azul (o RGB), no caso de telas. Uma foto é perfeitamente quantificável pois fazendo-se um *scanning* dela, transformando-se-a totalmente em números; posteriormente o computador pode imprimir-la com um resultado indistinguível do original. Com isso, pode-se estabelecer o que os computadores podem ou não podem processar: no primeiro caso, o que pode ser representado por números sem perda da qualidade original, e no segundo caso o que não pode ser representado por números, isto é, quando na quantificação se perde algo do original. Obviamente, seres vivos não podem ser quantificados, como também não o podem ser as sensações e sentimentos humanos pois, como foi visto, eles não são objetivos e universais. Portanto, inteligências humanas baseadas em sensações e sentimentos jamais poderão ser exercidas por máquinas. Jamais haverá uma inteligência humana artificial.

Achar-se que os robôs terão um dia todas as capacidades humanas é um engano produzido pelas forças que querem enganar o ser humano.

Referências

Steiner, R. *A filosofia da liberdade. Fundamentos de uma cosmovisão moderna*. GA 4. Trad. J.F. Torunski e R.Y. Santos. São Paulo: Antroposófica, 2022.

Setzer, V.W. [2024] "Conceitos, ideias, sensações, sentimentos e a cognição". Acesso em 10/5/25: <https://www.ime.usp.br/~vwsetzer/conceito-cerebro.pdf>

Setzer, V.W. [2025]. "O computador como instrumento do cientificismo". Acesso em 10/5/25:

<https://www.ime.usp.br/~vwsetzer/apresentacoes/comp-cient.ppsx>

e o artigo original que redundou nessa apresentação, publicado em 1976 nos *Anais do Simpósio Anual da Academia de Ciências do Estado de São Paulo* (acesso em 1/7/23:

<https://www.ime.usp.br/~vwsetzer/comp-cient.html>

¹ Este artigo é uma grande extensão do homônimo publicado na seção ESPAÇO ABERTO do jornal *O Estado de São Paulo*, edição de 13/11/2024, p. A4. Ele continua sendo dirigido a um público geral.