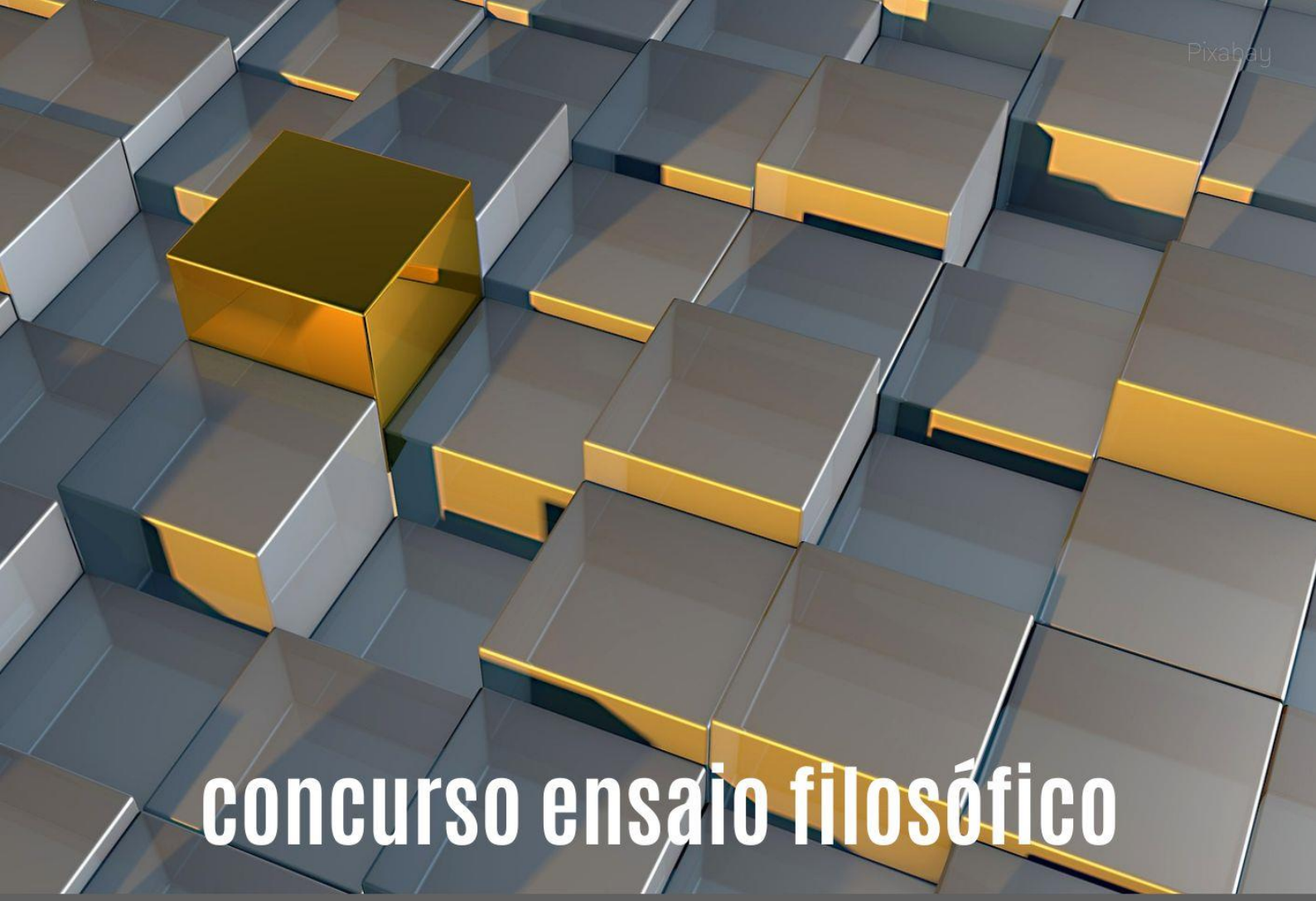

concurso ensaio filosófico

Pode a inteligência artificial conhecer? Uma análise sobre o estatuto epistémico da IA

Jaime Borralho

Prémio da 11.ª edição | 2024-2025

Associação de Professores de FILOSOFIA

REDE DE BIBLIOTECAS ESCOLARES

Ficha técnica

Título: Pode a inteligência artificial conhecer? Uma análise sobre o estatuto epistémico da IA

Autor: Jaime Borralho

Escola: Escola Secundária D. Manuel I - Agrupamento de Escolas 2 de Beja

Professora orientadora: Marcela Dantas da Silva

Associação de Professores de Filosofia em parceria com a Rede de Bibliotecas Escolares

Edição: Associação de Professores de Filosofia, Coimbra – 2025

Nota: o número final de páginas resultou na paginação introduzida pela APF, para efeitos de publicação.

Índice

Resumo

Os recentes avanços tecnológicos possibilitaram a criação de sistemas de inteligência artificial generativa de texto, o que tem levantado questões quanto às suas capacidades. O presente ensaio investiga o estatuto epistémico destes sistemas, questionando se podem realmente conhecer, tal como nós humanos. Para tal, procedeu-se a uma análise comparativa entre a forma como estes programas funcionam e a definição tradicional do conhecimento como crença verdadeira justificada. Argumenta-se que a inteligência artificial não pode ter conhecimento, para além de não constituir uma fonte confiável de conhecimento. A ausência de subjetividade e de modelos internos da realidade, aliada à dependência em induções estritamente estatísticas e o risco de interferências externas, impedem que estes sistemas satisfaçam os critérios clássicos do que significa conhecer. Adicionalmente, utiliza-se o argumento do quarto-chinês, proposto por John Searle, que fortalece a inexistência de uma verdadeira compreensão. Por fim, problematiza-se a sobrevalorização e a confiança excessiva que a sociedade contemporânea deposita nestas tecnologias, alertando-se para riscos e oportunidades, bem como para outras questões emergentes nesta área.

Palavras-chave: Inteligência Artificial, Inteligência Artificial Generativa de Texto, ChatGPT, Definição Tradicional do Conhecimento, Epistemologia

Descartes e as novas máquinas

Descartes (1637), citado por Cottingham (1986, p. 148), refletiu sobre as capacidades das máquinas, defendendo que “poderemos certamente conceber uma máquina construída para proferir palavras (…), mas não é concebível que uma tal máquina produza diferentes arranjos de palavras que deem resposta apropriadamente significante àquilo que seja dito em sua presença”.

Na época de Descartes, esta posição era perfeitamente plausível e justificável. Afinal, nenhuma engenharia feita de parafusos, alavancas e roldanas poderia alguma vez discutir filosofia com o leitor (Rachels, 2009). Contudo, os avanços tecnológicos, com destaque para a computação científica, vieram contestar o filósofo.

Em 1936, Alan Turing, ao resolver um problema de lógica matemática, descreveu a ideia de um dispositivo capaz de manipular símbolos, seguindo regras específicas. Anos mais tarde, este mesmo matemático previu que eventualmente estas máquinas tornar-se-iam capazes de reproduzir o comportamento inteligente. Antecipando-se, Turing vê os problemas filosóficos levantados, e, em 1950, aborda uma questão que iria moldar o futuro da filosofia aplicada à computação: poderão as máquinas pensar? (Oliveira, 2019).

A previsão de que as máquinas iriam simular a inteligência estava correta e, no último meio século, o desenvolvimento exponencial da inteligência artificial (IA) permitiu a criação de sistemas inéditos. Outrora, pensou-se que a linguagem flexível era o que nos destacava enquanto humanos, mas modelos como o ChatGPT — a IA generativa de texto mais enaltecida — questionam as nossas opiniões. Para além disso, o sucesso entre o público é inegável: o ChatGPT tornou-se a aplicação de consumo com o crescimento mais rápido da história (Hu, 2023).

No entanto, tal não significa que a questão sobre a possibilidade de as máquinas pensarem tenha sido respondida, sendo alvo de debate até aos dias de hoje. Na tentativa de encontrar uma resposta, são geradas outras perguntas de igual pertinência, que abordam outros ramos da Filosofia. Destaca-se a epistemologia, que se debruça sobre o conhecimento, tentando resolver problemas sobre a natureza e a validade do mesmo (Priberam, s.d.). Neste ensaio filosófico propomo-nos a examinar se estas novas máquinas podem efetivamente ser “sujeitos epistémicos”, ou seja, entidades capazes de conhecer, tal como nós humanos. Poderemos confiar, tão cegamente, nas IA? A

inteligência artificial tem realmente conhecimento? Iremos defender que a IA, em particular os sistemas de IA generativa de texto, não são sujeitos epistémicos e, por conseguinte, não podem conhecer.

Antes de esclarecer conceitos quanto ao conhecimento e à inteligência artificial, é importante referir que todo o conteúdo do ensaio tem por base o estado das IA, mais concretamente das ferramentas generativas de texto, até à data de elaboração do mesmo (2025).

Inteligência Artificial e conhecimento

A inteligência artificial representa um conjunto de tecnologias digitais, que, dependendo do seu foco e com base em objetivos definidos por humanos (explícitos ou implícitos), são capazes de fazer previsões, gerar conteúdo, fazer recomendações, entre outras competências, de acordo com os dados de entrada que recebe, podendo influenciar ambientes físicos ou virtuais (União Europeia, *Regulamento 2024/1689*, 2024, artigo 3º).

No âmbito destas capacidades, a inteligência artificial generativa é aquela que tem como foco a geração de conteúdo, incluindo a criação de imagens, texto, áudio, vídeo, entre outros. Neste ensaio, vamos focar-nos somente na criação de texto, popularizada por modelos como o ChatGPT, o Copilot, ou o Gemini. Assim, a partir deste ponto o conceito de inteligência artificial será entendido como IA generativa de texto. Torna-se então necessário fazer uma breve análise sobre o funcionamento destes sistemas, de forma a possibilitar uma melhor compreensão dos argumentos adiante.

Como funcionam as Inteligências Artificiais generativas de texto?

O primeiro conceito fundamental no funcionamento da IA generativa de texto é o de “redes neuronais”. Na sua essência, uma rede neuronal é uma tentativa de representar digitalmente parte do funcionamento do cérebro humano. Estas redes são constituídas por camadas de neurónios artificiais, estando estes conectados por “pesos” (valores ajustáveis que alteram a influência dos neurónios na rede). Desta forma, cada neurónio realiza equações matemáticas e as "perguntas" (*inputs*) percorrem a rede até chegar às "respostas" (*outputs*). Ao enviarmos uma mensagem a um destes programas, o sistema começa por converter as palavras em *tokens*. Neste contexto, *tokens* são unidades de informação, como palavras ou fragmentos delas, que estão associados a um determinado número. Podemos dizer que estes sistemas atribuem a cada palavra do dicionário um "número de identificação". Por sua vez, estas sequências numéricas

são transferidas para a rede neuronal, mas passam ainda por uma transformação adicional. A cada palavra é atribuído um *embedding*, isto é, uma representação da essência da palavra, sob a forma de um vetor, num espaço multidimensional. Pode parecer complexo, mas no fundo é uma maneira de mapear as palavras num "espaço de significado", onde palavras com semânticas semelhantes são posicionadas próximas neste espaço (Wolfram, 2023).

Isto é importante para compreendermos como é que estas IA funcionam baseando-se em probabilidades. Tendo em conta o contexto, elas geram respostas ao escrever a palavra que, com base nas anteriores, tem maior probabilidade de aparecer. É como se se estivesse a perguntar constantemente "qual a palavra mais provável de aparecer depois desta?". Estas probabilidades são determinadas no treino da IA, onde o programa é alimentado com enormes quantidades de páginas, artigos e livros, disponíveis na web. Estes dados têm a importante função de "ensinar" ao programa padrões linguísticos, como estruturas gramaticais e significados (o que também ajusta os pesos e permite os *embeddings*). Com estas probabilidades, o nosso modelo escolhe a palavra mais provável e converte-a de volta num *token*, continuando esse processo até gerar texto coerente e fluido, *aparentando* ter conhecimento (Wolfram, 2023).

O que é conhecimento?

Sabemos que as IA em questão criam texto no formato de resposta. Por isso, neste ensaio, vamos abordar exclusivamente o conhecimento proposicional, ou seja, o conhecimento que, através de frases declarativas, exprime algo que é verdadeiro ou falso (Sober, 2008). Com isto em mente, sempre que nos referirmos a conhecimento, assumiremos o conhecimento proposicional, o “saber-que”.

A filosofia clássica apresenta uma definição de conhecimento proposicional. No diálogo de Platão *Teeteto* (369 a.C.), são discutidas três condições necessárias, que apenas em conjunto são suficientes para verificar a existência de conhecimento. Estas condições formam a definição tradicional do conhecimento (DTC).

Comecemos pela condição de crença. Na filosofia, uma crença é um estado mental — ou melhor, como iremos aprofundar mais adiante, uma atitude mental. Ter uma crença significa acreditar numa proposição, seja a proposição verdadeira ou falsa. Quando um sujeito afirma "eu sei P", tal significa "eu acredito que P". Assim, para existir conhecimento, de acordo com a DTC, o sujeito precisa de crer no que afirma saber (Pritchard, 2006; Sober, 2008). Na nossa análise, o sujeito é a IA generativa de texto.

Do lado objetivo temos a condição de verdade. Para existir conhecimento não basta acreditar numa proposição com qualquer valor. É necessário que esta seja

verdadeira, o que significa que a proposição precisa de corresponder à realidade — ser factual — independentemente de como é a mente de cada um. O sujeito tem então de acreditar em algo verdadeiro (Pritchard, 2006; Sober, 2008). Mais adiante iremos discutir se a IA respeita e compreende este critério.

Por fim, a crença verdadeira tem de estar justificada. O conhecimento exige dados de apoio, boas razões, que o suportem. A justificação assegura que o conhecimento não é aleatório ou acidental (Sober, 2008). Por exemplo, caso eu acerte nos números da lotaria, não é legítimo afirmar "eu sabia que eram estes os números", já que não existe nenhum sustento lógico para esta crença, que se veio a tornar verdadeira. No que toca às IA, vamos notar que as suas justificações diferem das humanas, todavia, tal não deve levar-nos a assumir de imediato que carecem de justificações válidas.

Como podemos notar, a DTC oferece um bom ponto de partida para avaliar a possibilidade de conhecimento em modelos como o ChatGPT. Mesmo assim, devido à sua natureza filosófica, não existe uma única abordagem para a definição de conhecimento, existindo objeções à DTC. Por isso, de forma a ultrapassar esta dificuldade inevitável, procurámos negar todas as condições, tendo também em conta que as condições “crença” e “verdade” são amplamente reconhecidas (Pritchard, 2006).

Passemos então à análise comparativa entre as respostas da IA e a DTC.

Comparação e análise

Têm crenças?

Como já referimos, crenças são estados mentais. Para existir um estado mental como o de acreditar, é necessário que haja algum tipo de subjetividade. Enquanto sujeitos epistémicos, cabe-nos avaliar, refletir e pensar sobre proposições, quer recorrendo à razão, quer à experiência. É por este motivo que as crenças não são consensuais, mas sim, subjetivas. A crença implica intenção e compreensão, o que resulta numa experiência pessoal sobre o que se deve acreditar. Aliás, ao argumentar que crenças são estados mentais que requerem subjetividade, podemos até enquadrá-las como "atitudes mentais", como são normalmente designadas.

Com este raciocínio em mente, vamos então aplicá-lo ao objeto de estudo do nosso ensaio. As IA generativas de texto têm subjetividade? Dependendo da resposta, conseguiremos verificar se podem ou não ter crenças.

Definimos, implicitamente, subjetividade como a capacidade de ter uma experiência interna pessoal. Ao analisarmos a existência de subjetividade no ChatGPT, deparamo-nos com algo: a IA *aparenta* ter subjetividade. Ao interagir, a resposta chega a incluir *emojis* e sugestões como "se quiser, posso aprofundar o tópico". Contudo, esta

interação carinhosa é mero reflexo de equações complexas. Os dados de treino "instruíram" a máquina a comunicar desta forma. Não existe aqui nenhuma compreensão ou questionamento, apenas matemática.

Podemos inclusive fazer uma comparação entre estas IA com uma calculadora comum. Ambas processam informação com uma linguagem diferente da comunicada; a IA com *tokens*, a calculadora com código binário. Também recebem, por humanos, *inputs* que são analisados e resultam em *outputs*. E por fim, ambas utilizam razões matemáticas para este processamento (Woodford, 2024). Mas porque é que uma calculadora parece tão impessoal e sem uma experiência interna? A resposta é simples: uma calculadora é transparente e previsível. É limitada pela função de calcular. Do outro lado temos a IA, que, estando alojada em servidores poderosíssimos, utiliza milhões de equações que incluem álgebra linear, cálculo vetorial e probabilístico (Gutierrez, 2024). Ao "possuir" um domínio linguístico que pode ser contextualizado, e que absorveu milhares de interações humanas durante o treino, a IA consegue mascarar a sua falta de subjetividade, contrariamente à calculadora.

Com isto, podemos concluir que a inteligência artificial não tem subjetividade, o que a impede de ter atitudes mentais. Ora, sem atitudes mentais, a IA não tem crenças, assim como não tem qualquer emoção, sentimento, desejo, etc. — aquilo que os filósofos denominam de "*qualia*".

Perante a posição defendida, é possível fazer uma objeção recorrendo às teses funcionalistas da mente. Afinal, se temos um sistema que copia fielmente atitudes mentais, que passa em qualquer teste para distinguir seres subjetivos de não subjetivos, então porque haveríamos de recusar-lhe esse estatuto? Quando um ser humano apresenta comportamentos linguísticos adequados, assumimos subjetividade. Porquê exigir mais da IA? Esta objeção parece convincente, mas falha após análise. Como já referimos, é tentador argumentar que o ChatGPT tem subjetividade, esquecendo as consequências desse estatuto. Na hipótese de a IA funcionar como um ser subjetivo, tal implicaria uma capacidade de ir para além das ideias já concebidas. O problema é que não encontramos nada disso no seu funcionamento, muito menos nas suas respostas.

Mas esta implicação não é apenas metafísica. Podemos utilizar a evolução dos paradigmas na física para ver na prática quais são os efeitos da subjetividade no conhecimento. Inicialmente, Aristóteles defendeu que, ao largar uma maçã, esta cai no chão já que esse é o seu "sítio natural". Séculos mais tarde, Newton sugere uma força invisível chamada "gravidade" para explicar este fenómeno. Mais adiante, surge Einstein, redefinindo o conceito de gravidade como o resultado da curvatura do espaço-tempo, provocada pela massa da Terra. Estas mudanças ocorreram porque a subjetividade implica, numa forma abstrata, criatividade, inovação (Chomsky et al.,

2023). A capacidade de conceber ideias que nunca ninguém tinha sequer idealizado, rompendo pressupostos. Em oposição, a IA limita-se a seguir estatísticas derivadas dos dados de treino, de forma dogmática. Por mais surpreendente que a IA possa parecer, esta capacidade de transcender as premissas só é, até à data, observável em seres biológicos. O presente raciocínio encaixa-se, por sinal, na inferência válida "modus tollens": se a IA tem subjetividade então é capaz de refletir sobre proposições (o que resulta em criatividade). Tal não se verifica. Logo, a IA não tem subjetividade. Mais uma vez, sem subjetividade não há crenças — nega-se a primeira condição da DTC.

Sabem verdades?

Avancemos para o próximo critério. É na condição "verdade" que a IA compromete, de forma mais notória, não só o seu estatuto enquanto sujeito epistémico, como também a confiança que lhe é dada atualmente.

Em primeiro lugar, sabemos que a IA comete erros, podendo fornecer informações enganosas. Como não tem um verdadeiro contacto com a realidade é expetável que incorra em falácias, contradições, informações falsas ou até mesmo factos inventados — as chamadas "alucinações da IA". Mas de que forma é que isto afeta o seu estatuto epistémico? Afinal, um ser humano, que é capaz de conhecer, também erra.

A diferença reside no facto de a IA não compreender o que é verdade, algo que o conhecimento necessita. Enquanto sujeitos epistémicos, nós vamos conhecendo o mundo e construímos, mesmo que inconscientemente, modelos da realidade. Estes modelos são, em boa parte, compartilhados entre sujeitos. Mas ao analisar a IA não notamos qualquer modelo da realidade, e é exatamente esta ausência que a impossibilita de distinguir o verdadeiro do falso. Apenas consegue gerar respostas que soam plausíveis — e que até podem estar corretas — baseando-se na imensa base de dados que tem (Caldas, 2023). A IA é programada para seguir probabilidades e não para garantir a veracidade da informação. Apesar disso, estes sistemas apresentam um impressionante “domínio” da língua natural: frases bem estruturadas e objetivas, coerência textual, vocabulário diversificado, etc. Tudo isto dá clareza às respostas, resultando em textos geralmente convincentes, mesmo com o conteúdo incorreto.

Neste sentido, alimentar uma IA com artigos de física não a torna conhecedora da matéria. Isso apenas lhe dá a capacidade de reproduzir padrões linguísticos desses textos, como, por exemplo, a terminologia utilizada. A "verdade", para a IA, depende da compreensão de *outros* sujeitos epistémicos, algo que não a torna conhecedora de facto, como tantas vezes é considerada.

Como se não bastasse, esta dependência de outros sujeitos resulta frequentemente em algoritmos preconceituosos, como é mostrado no documentário *Coded Bias* (Kantayya, 2020). Na realidade, o processo de "ensinar" uma máquina tende a ser influenciado pelos pressupostos e preconceitos dos programadores. Tal resulta num programa fundamentalmente enviesado. É importante notar que os humanos também podem ser preconceituosos, mas estes têm a capacidade de refletir e de alterar as suas ideias, contrariamente à IA que fica "forçada" a aceitar os dados de treino com que é alimentada, estejam eles corretos ou não.

Conseguem justificar?

A condição "verdade" abre terreno para prosseguir a nossa análise, agora no critério de justificação. Antes de mais, sabemos que para um humano justificar o conhecimento pode recorrer a diversas fontes, nomeadamente, aos seus sentidos (como a visão), ao testemunho de fontes confiáveis (como um livro) ou ao raciocínio dedutivo (como razões lógicas). Então como poderá uma IA justificar? Para responder a esta questão, precisamos de analisar o fundamento do raciocínio utilizado em sistemas como o ChatGPT.

Do ponto de vista técnico, a justificação da IA assenta em modelos estatísticos, que, numa conversão à linguagem epistemológica, enquadram-se no raciocínio indutivo. Neste tipo de raciocínio os casos particulares nas premissas são utilizados para generalizar ou prever. Assim, uma inferência indutiva seria, por exemplo, "até agora só vi cisnes brancos, logo todos os cisnes são brancos". Quanto maior a amostra analisada, mais forte o argumento se torna. Mesmo assim, estes argumentos têm sempre por base a probabilidade, nunca revelando verdades necessárias. De forma semelhante, a IA aprende por indução, no processo denominado "*machine learning*" (Arão, 2024).

Curiosamente, o *machine learning* assemelha-se ao empirismo: começa-se com uma "tábua rasa" que vai sendo preenchida pela experiência — neste caso, a experiência são os dados com que a IA é alimentada. Note-se que foi só a partir do século XXI, com o aumento do poder computacional, o desenvolvimento da Internet e a consequente disponibilização massiva de dados, que foi possível programar um *software* que consiga ter "experiências" suficientemente significativas para "*agir"* como um humano. No século passado eram utilizados, em grande parte, raciocínios dedutivos que, apesar de se terem mostrado fantásticos em jogos como o xadrez, não mostraram grande potencial prático. Contrariamente ao xadrez que tem regras bem definidas e um espaço limitado, o mundo real é incerto e variável. Por isso, foi necessário desenvolver outras abordagens, nas quais os programas extraem padrões contidos nos dados de

treino, através de raciocínios indutivos. As máquinas passam agora a generalizar informação, o que cria as respostas estatisticamente plausíveis que temos vindo a comentar. No fundo, nestes sistemas a realidade é derivada estatisticamente, a partir de semelhanças históricas, presentes nos dados (Oliveira, 2019; Oliveira, 2025).

Mas será esta justificação suficiente para considerarmos que a IA tem justificações válidas? Não. Na verdade, esta conceção é meramente funcional, já que o que se cria são programas que apenas generalizam informação de forma acrítica. Cria-se uma ferramenta que tenta justificar-se utilizando dados que não compreende. Ao colocarmos um humano a justificar o conhecimento nesta lógica, então fica evidente a ausência de uma justificação apropriada por parte da IA. Tal como escreveu Cristian Arão (2024, p. 14): "espantamo-nos com a capacidade da inteligência artificial de copiar o estilo de algum artista, mas não nos atentamos para o facto de que isso é o máximo que ela pode fazer". Adicionalmente, este excesso de confiança em induções estatísticas confronta a IA com outros problemas fundamentais na epistemologia, como os levantados por David Hume — filósofo que defendeu que a indução carece de justificação epistémica.

No fundo, o recurso à estatística e à indução na tecnologia, já levanta, por si só, dúvidas e incertezas. Ora, soma-se a isto um sistema que não possui uma compreensão da realidade e obtemos o que chamamos de “inteligência artificial generativa”.

Posto isto, a posição defendida ao longo do ensaio torna claro um ponto crucial que assombra a IA: a falta de uma verdadeira compreensão. Portanto, na tentativa de fortalecer o problema e as posições aqui sustentadas, bem como para finalizar a nossa argumentação, passamos a apresentar o argumento do quarto-chinês, de John Searle.

A experiência mental de John Searle

Em 1984, o filósofo americano John Searle publicou a obra "Mente, Cérebro e Ciência". Nesta obra, que recai essencialmente sobre a Filosofia da Mente aplicada à computação, Searle propõe uma experiência mental cujas implicações se sentem mesmo após 40 anos do seu lançamento.

Então, em que consiste esta experiência? Imaginemos um homem que desconhece por completo a *língua chinesa* (o mandarim) e que está fechado num quarto com um livro de instruções imenso (com milhões de páginas). Este quarto tem apenas uma porta com uma abertura, pela qual se conseguem passar pedaços de papel. Ocasionalmente, pode ser enviado por esta abertura um pedaço de papel, com caracteres em chinês — que para o homem não passam de rabiscos sem qualquer significado. Contudo, foi pedido ao homem que, quando entrasse um papel, ele deveria abrir o livro de instruções e procurar uma resposta para enviar em retorno, noutro papel.

Neste livro, estão presentes instruções elaboradas em português que indicam o que o homem deve responder ao receber as mensagens em chinês. Por exemplo, uma instrução deste livro seria: "ao ver o símbolo w e x, responda com os símbolos y e z". Desta forma, o homem passa horas a receber mensagens, a analisar o livro, e a enviar de volta outras mensagens, que para ele nada significam (Searle, 2000; Rachels, 2009).

O que o homem não sabe é que do outro lado da abertura está uma mulher de Pequim, a responsável pelos pedaços de papel que ele recebe. A mulher envia o papel, aguarda, e recebe de volta uma resposta. Nesta troca de mensagens, a mulher acredita que acabou de discutir assuntos como política e literatura. Ela acredita que no outro lado existe um fluente em chinês, mas sabemos que isso não é verdade. No outro lado está apenas um homem que, cegamente, reproduz símbolos de um livro enorme (transforma *inputs* em *outputs*) (Searle, 2000; Rachels, 2009).

Vamos agora aplicar esta experiência mental ao nosso ensaio. Podemos afirmar que a mulher, fora do quarto, acredita que o homem compreende e sabe aquilo que escreve. Apesar de não o ver pessoalmente, *presume* que, do outro lado, existe um ser capaz de conhecer e, inclusive, de transmitir conhecimento confiável. Afinal, as mensagens coerentes que recebe dão-lhe fortes motivos para o fazer. Contudo, como sabemos, as suas convicções, mesmo que sensatas, não correspondem à realidade. De modo semelhante, a grande maioria dos utilizadores da IA generativa de texto considera estes sistemas uma fonte de conhecimento e trata-os, mesmo que inconscientemente, como sujeitos que possuem realmente conhecimento. No entanto, a nossa análise mostrou que para existir conhecimento não basta manipular símbolos: é preciso compreender. Com isto, é curioso observar que os avanços tecnológicos foram capazes de criar “quartos-chineses” tão eficazes que conseguem enganar não só uma pessoa, mas praticamente todo o mundo.

Iludidos pela inovação

A partir dos argumentos desenvolvidos, podemos concluir que sistemas de inteligência artificial generativa de texto não podem ter conhecimento, não podem “saber-que”. Conhecer requer, mesmo que implicitamente, compreensão — para além de crença, de verdade e de justificação, algo que a IA não consegue ter. Por conseguinte, podemos até afirmar que estes programas, em termos filosóficos, têm o mesmo estatuto epistémico que um papagaio "que fala" — ambos interagem, mas não percebem o que estão a transmitir. Pode parecer um insulto à IA, dada a sua popularidade, mas esta é, numa escala diferente, a sua essência.

Existe uma sobrevalorização destes sistemas, derivada, em grande parte, do desconhecimento da sua verdadeira natureza. Por este motivo, o presente ensaio pode ser encarado como uma afronta à popularidade da IA generativa, na medida em que se desafia não só a capacidade de conhecer da IA, como também a sua legitimidade enquanto fonte confiável de conhecimento. Esta posição sustenta-se na ausência de subjetividade e de uma capacidade criativa, na inexistência de modelos da realidade, bem como nos riscos de parcialidade e na dependência em bases estatísticas. Retomando as palavras de Arão (2024, p. 12): "a exatidão dos números só existe na imaterialidade do reino da matemática pura (…) toda e qualquer aplicação no mundo material sofrerá com imprecisões, incertezas". Cabe, então, à Filosofia desvendar estes problemas, que, neste caso, aplicam-se mais à computação e à matemática.

Por outro lado, a nossa análise-reflexiva impulsiona outros problemas noutros domínios. Por exemplo, será que devemos atribuir, eventualmente, direitos e obrigações a estes sistemas? Têm estatuto moral? Que acesso é que devem ter a dados pessoais? E quem é o responsável quando a IA for usada para espalhar desinformação? Estas são algumas das questões emergentes, às quais importa responder. É certo que todos podemos beneficiar da IA, mas cabe reconhecer as suas limitações ou ameaças, que podem, inclusive, igualar as suas oportunidades. Acreditamos que o mais sensato é informar e sensibilizar a população, tanto das oportunidades como dos desafios destas tecnologias. No contexto escolar, por exemplo, a utilização destes programas não deve ser um "tabu", mas antes merece regulamentação e debate.

Na realidade, o verdadeiro problema não é a inteligência artificial em si, mas antes a falta de discernimento humano para guiá-la com sabedoria. Mas deixemos esse assunto para uma outra discussão.

Referências bibliográficas

Arão, C. (2024). Por trás da inteligência artificial: uma análise das bases epistemológicas do aprendizado de máquina. *Trans/Form/Ação, 47*(3), 1-17. https://doi.org/10.1590/0101-3173.2024.v47.n3.e02400163

Caldas, B. (2023, 11 de maio). A verdade do ChatGPT. *Observador*. https://observador.pt/opiniao/a-verdade-do-chatgpt/

Chomsky, N., Roberts, I., & Watumull, J. (2023, 8 de março). The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Cottingham, J. (1986). *A filosofia de Descartes.* Edições 70.

Gutierrez, D. (2024, 12 de novembro). Why mathematics is essential for data science and machine learning. *InsideAI News*. https://insideainews.com/2024/11/12/why-mathematics-is-essential-for-data-science-and-machine-learning/

Hu, K. (2023, 2 de fevereiro). ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Kantayya, S. (Realizadora). (2020). *Coded Bias* [Documentário]. Netflix. https://www.imdb.com/pt/title/tt11394170/

Oliveira, A. (2019). *Inteligência artificial.* Fundação Francisco Manuel dos Santos.

Oliveira, A. (2025). *A inteligência artificial generativa.* Fundação Francisco Manuel dos Santos.

Priberam. (s.d.). *Epistemologia*. Dicionário Priberam da Língua Portuguesa. https://dicionario.priberam.org/epistemologia

Pritchard, D. (2006). *What is this thing called knowledge?* Taylor & Francis Group. https://doi.org/10.4324/9780203968468

Rachels, J. (2009). *Problemas da filosofia.* Gradiva.

Searle, J. (2000). *Mente, cérebro e ciência*. Edições 70.

Sober, E. (2008). *Core questions in philosophy.* Prentice Hall.

União Europeia. (2024). Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria da inteligência artificial. http://data.europa.eu/eli/reg/2024/1689/oj

Wolfram, S. (2023). *O que faz o ChatGPT e como funciona.* Casa das Letras.

Woodford, C. (2024, 15 de fevereiro). Calculators. *Explain that Stuff*. https://www.explainthatstuff.com/calculators.html