



VIRTUDES E VÍCIOS EPISTÊMICOS E O CHATGPT *

Epistemic virtues and vices and ChatGPT

Veronica de Souza Campos **

Resumo: Este artigo analisa o ChatGPT sob a perspectiva da epistemologia das virtudes, argumentando que, apesar de certos benefícios, a ferramenta é mais propensa a fomentar vícios epistêmicos do que a cultivar virtudes. Após delinear as noções centrais da epistemologia das virtudes, particularmente a distinção entre virtudes e vícios intelectuais, o artigo avalia o impacto epistêmico de modelos de linguagem de grande porte (LLMs) na ciência e na educação. Embora o ChatGPT possa oferecer vantagens práticas como velocidade, acessibilidade e suporte pedagógico, ele também incentiva a passividade intelectual, enfraquece a memória, a autoria e o pensamento crítico, e reproduz vieses sociais. Essas dinâmicas geram o que estudos recentes descrevem como “dívida cognitiva”, uma erosão progressiva da autonomia e da responsabilidade intelectual. O artigo destaca que delegar tarefas epistêmicas ao ChatGPT frequentemente distorce propósitos, privilegiando resultados em detrimento de processos essenciais ao florescimento intelectual. Conclui que o uso generalizado e acrítico de LLMs corre o risco de consolidar a preguiça, a indiferença epistêmica e a desonestidade, minando, em vez de fortalecer, nossas capacidades como agentes epistêmicos.

Palavras-chave: Epistemologia baseada em virtudes. ChatGPT. Vícios intelectuais.

* Artigo recebido em 05.09.2025 e aprovado para publicação em 20.11.2025.

** Professora substituta no Departamento de Filosofia da Universidade Federal de Minas Gerais (UFMG) e pesquisadora de pós-doutorado na Faculdade Jesuíta de Filosofia e Teologia (FAJE). Av. Dr. Cristiano Guimarães, 2127 — Planalto, Belo Horizonte — MG, 31720-300. ORCID: <https://orcid.org/0000-0001-6242-5600>. E-mail: 182vkai@gmail.com

Abstract: This article analyses ChatGPT from the perspective of virtue epistemology, arguing that, despite certain benefits, the tool is more prone to fostering epistemic vices than cultivating virtues. After outlining the central notions of virtue epistemology, particularly the distinction between intellectual virtues and vices, the paper assesses the epistemic impact of large language models (LLMs) in science and education. While ChatGPT may offer practical advantages such as speed, accessibility, and pedagogical support, it also encourages intellectual passivity, weakens memory, authorship, and critical thinking, and reproduces social biases. These dynamics generate what recent studies describe as “cognitive debt,” a progressive erosion of intellectual autonomy and responsibility. The article highlights that delegating epistemic tasks to ChatGPT often distorts purposes, privileging results over processes essential to intellectual flourishing. It concludes that the widespread and uncritical use of LLMs risks consolidating laziness, epistemic insouciance, and dishonesty, thereby undermining rather than enhancing our capacities as epistemic agents.

Keywords: Virtue-based epistemology. ChatGPT. Intellectual vices.

Introdução

O ChatGPT é uma inteligência artificial do tipo Modelos de Linguagem de Grande Porte (em inglês *Large Language Model*, tipicamente referida por meio da sigla LLM), muito comentada atualmente por ser capaz de interagir e gerar linguagem humana.¹ Como tal, ele simula processos de inteligência humana, tendo sido “projetado para tarefas de processamento, geração e compreensão de linguagem natural”.² O ChatGPT não é a única aplicação de um LLM, embora seja um dos mais conhecidos e discutidos (outras ferramentas da mesma natureza incluem o Gemini, da Google, e o Dolly, da Databricks). Todos eles funcionam emulando a conversa humana, na medida em que interagem com os usuários principalmente por meio de texto: ao receber uma entrada, que geralmente é composta de uma solicitação feita pelo usuário em linguagem natural, o Chat GPT analisa os padrões comparando-os com os dados com os quais foi treinado e, em seguida, usa os mesmos padrões para gerar e emitir respostas. A saída produzida pelos chatbots é tipicamente um bloco de texto que parece ter sido escrito por uma pessoa.

¹ RAY, Partha. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, in *Internet of Things and Cyber-Physical Systems*, (3):1, 2023, p. 121.

² RAY, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, p. 121-122.

Desde o lançamento do ChatGPT, em 2022, os LLM passaram rapidamente a fazer parte de nossa vida cotidiana, tendo sido adotados em nosso trabalho, estudo e vida pessoal, onde supostamente ajudam a acelerar pesquisas, melhorar peças de texto e criar conteúdo, dentre outras funções, em um ritmo sem precedentes. O entusiasmo por essas ferramentas teve uma taxa de adoção ainda maior do que a da internet ou dos computadores. Como Alier, García-Peñalvo & Camba³ discutem, essa tecnologia passou de desconhecida para central na cultura digital em tempo recorde, tendo atingido um milhão de usuários em cinco dias e 100 milhões em menos de dois meses. Como acontece com toda nova tecnologia, o ChatGPT pode impulsionar a sociedade de diversas maneiras, mas também é de se esperar que seu uso traga prejuízos.

Uma pergunta que deve estar em nosso horizonte é: na medida em que recorreremos cada vez mais ao ChatGPT para realizar tarefas que de outro modo realizaríamos nós mesmos, qual o impacto que essa dependência pode ter em termos epistêmicos, i.e., em termos de como nos relacionamos com o conhecimento? Será que o uso recorrente e intensivo pode afetar nossa própria inteligência? Uma forma de desempacotar essa preocupação é formulá-la em termos de virtudes e vícios epistêmicos. Pode o ChatGPT contribuir para o desenvolvimento de virtudes epistêmicas (traços como humildade, curiosidade e mente aberta, etc.), ou será que ele é uma ferramenta muito mais propensa a contribuir para a proliferação de vícios do intelecto, como a preguiça?

Com esse questionamento em mente, o artigo procederá da seguinte forma. Na seção 1, farei uma apresentação panorâmica da epistemologia baseada em virtudes (*virtue-based epistemology*) e de suas noções centrais – noções como virtude, vício, caráter, etc. Na seção 2, apresento um levantamento sobre alguns dos bônus (isto é, ganhos epistêmicos) associados ao uso de LLMs em algumas de nossas principais atividades distintamente epistêmicas, a saber, a ciência e a educação. Na seção 3, discuto alguns dos principais ônus epistêmicos oriundos desse uso. Na seção 4, discuto um problema relacionado, de distorção de propósitos. Concluo, na última seção, resumizando o que foi discutido.

1. Epistemologia baseada em virtudes

A epistemologia baseada em virtudes é uma corrente em epistemologia contemporânea composta por um conjunto de abordagens que compartilham dois pressupostos fundamentais: (i) a epistemologia deve ser concebida como

³ ALIER, Marc, GARCÍA-PEÑALVO, Francisco; CAMBA, Jorge. Generative Artificial Intelligence in Education: From Deceptive to Disruptive, in *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 2024, p. 5-14.

uma disciplina normativa; e (ii) a avaliação epistêmica deve ter como principal objeto os agentes cognitivos, e não meramente os estados doxásticos por eles mantidos. Em outras palavras: importa, a esse ramo da epistemologia, examinar os indivíduos e seu desempenho epistêmico (o quão bem – ou o quão mal – eles se saem em tarefas epistêmicas que interessam, como a pesquisa, o ensino, a aprendizagem, dentre outras, e porquê).

A proposta de uma epistemologia ancorada na noção de virtude foi inicialmente elaborada por Ernest Sosa⁴ e, desde então, desenvolveu-se em diferentes direções, notadamente a partir da distinção entre duas perspectivas. A primeira delas compreende as virtudes intelectuais como competências ou habilidades cognitivas, e ficou conhecida como *confiabilismo da virtude*. A segunda perspectiva, que ficou mais conhecida a partir do trabalho de Linda Zagzebski,⁵ entende virtudes intelectuais como disposições de caráter e recebeu o nome *responsabilismo da virtude*. Algumas virtudes parecem ser mais bem classificadas como habilidades cognitivas, à maneira de Sosa – como memória acurada ou percepção visual refinada – ao passo que outras são mais adequadamente descritas em termos de traços de caráter, à maneira de Zagzebski, como é o caso da humildade e da curiosidade.

Na taxonomia proposta por Jason Baehr,⁶ diferentes desenvolvimentos da epistemologia baseada na virtude podem ser classificados em duas categorias principais, de acordo com sua relação com a epistemologia tradicional (abordagens conservadoras e autônomas), e ainda subdivididas em duas modalidades quanto ao alcance explicativo conferido às noções de virtude e de vício (abordagens fortes e fracas). As abordagens conservadoras recorrem ao conceito de virtude intelectual como instrumento para tratar problemas clássicos da epistemologia. Para seus defensores, as noções aretaicas fornecem recursos conceituais que auxiliam na resolução desses problemas, o que implica que tal vertente mantém como núcleo de interesse as questões tradicionalmente centrais da disciplina, tais como: O que é o conhecimento? É possível conhecer? Qual a diferença entre opinião e ciência? Entre outras. As abordagens autônomas, em contraste, privilegiam o exame das virtudes intelectuais em si mesmas, de modo relativamente independente das questões legadas pela epistemologia clássica, ainda que conservem uma orientação epistêmica ampla.

Tanto as abordagens conservadoras quanto as autônomas podem assumir a forma fraca ou forte. Abordagens fracas, sejam do tipo conservador ou autônomo, sustentam que recorrer às noções de virtudes e vícios intelec-

⁴ SOSA, Ernest. *The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge*. Midwest Studies in Philosophy, v. 5, 1980, p. 3-25.

⁵ ZAGZEBSKI, Linda. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press, 1996.

⁶ BAEHR, Jason. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford, Oxford University Press, 2012, p. 12.

tuais é um artifício para complementar, mas não substituir, as categorias clássicas no enfrentamento dos problemas epistemológicos. Abordagens fortes, por sua vez, defendem que as noções de virtudes e vícios intelectuais devem substituir integralmente os quadros explicativos herdados da epistemologia tradicional. Nesse sentido, o exame conduzido no presente trabalho assume uma perspectiva autônoma fraca: pretendemos examinar se virtudes epistêmicas podem ser cultivadas e desenvolvidas no indivíduo a partir de experiências com o uso de uma certa ferramenta (o ChatGPT), ou se essa ferramenta tem efeitos muito mais deletérios em termos de o quão virtuoso seu usuário tende a se tornar, isto é, se a ferramenta contribui para a instalação de vícios epistêmicos. Através esse exame não pretendemos engajar com nenhum dos problemas da epistemologia clássica e nem substituir análises já apresentadas. Ao contrário, pretendemos apenas ver até que ponto é possível recomendar ou não recomendar o uso do ChatGPT com base em benefícios e prejuízos que ele traz a nós enquanto agentes epistêmicos – ver se com ele nós nos tornamos agentes epistêmicos melhores do que costumávamos ser antes dele.

Agora, o que exatamente são virtudes e vícios intelectuais? Tipicamente, virtudes intelectuais são definidas como excelências relacionadas ao conhecimento e à verdade, isto é, à sua aquisição, transmissão ou manutenção. Essas excelências podem ser habilidades (no entendimento confiabilista, descrito acima) ou traços de caráter (no entendimento responsabilista). Num caso como no outro, trata-se de características que estimulam uma lida adequada com o conhecimento e com a verdade, pois melhoram a capacidade ou disposição de alguém para investigar apropriadamente, para entender fatos e processos, para retê-los na memória e explicá-los a outros. Exemplos de virtudes intelectuais incluem curiosidade, honestidade intelectual, humildade, diligência, etc. Os vícios, por sua vez, são características que prejudicam, em vez de melhorar, essa capacidade ou disposição. Isto é, são falhas. Exemplos incluem desonestidade intelectual, preguiça, ingenuidade, arrogância, dentre outros.

Conforme observa Eric Kraemer,⁷ os teóricos da epistemologia baseada na virtude têm concentrado maior atenção nas virtudes intelectuais do que nos vícios correspondentes. No âmbito responsabilista, tanto a epistemologia da virtude quanto a epistemologia do vício compartilham a premissa de que a cada virtude correspondem um ou mais vícios como seu contraponto. Em linhas gerais, a virtude costuma ser concebida como um meio-termo entre dois extremos, sendo estes caracterizados precisamente como vícios (a exemplo da coragem, que é um meio-termo entre o vício da covardia, por um lado, e o da temeridade, por outro; e da generosidade, que é um meio-termo entre avareza, por um lado, e a liberalidade, por outro). Faz

⁷ KRAEMER, Eric. Virtue Epistemologies and Epistemic Vice, in *Athens Journal of Humanities & Arts*, v.2, 2015, p. 19-30.

parte da noção de virtude, tal como entendida dentro desse framework teórico, que sua aquisição é algo que se consegue por meio do hábito, enquanto vícios se proliferam diante da falta de cultivo das virtudes.

Quando temos vícios de caráter, nosso desempenho ao descobrir, reter e transmitir a verdade é piorado. Por exemplo, o vício da ingenuidade torna o indivíduo propenso a acreditar em proposições improváveis para as quais não há evidências.⁸ A indiferença epistêmica (*epistemic insouciance*) leva a pessoa a não se importar com se algo é verdadeiro ou falso e consequentemente a não investigar assuntos que precisam ser investigados, ou a investigar de forma frouxa ou imprudente.⁹ A injustiça epistêmica prejudica a credibilidade de uma terceira pessoa, tornando mais difícil a partilha do conhecimento.¹⁰ Quando alguém tem um desempenho epistêmico ruim, isto é, quando a pessoa fracassa em obter, reter e transmitir conhecimento, esses traços desempenham um papel significativo na explicação de como e por que ela fracassou.

Os epistemólogos da virtude discordam entre si sobre o que, exatamente, torna uma virtude virtuosa e um vício vicioso. Enquanto autores confiabilistas como Swank,¹¹ Cassam¹² e Campos¹³ afirmam que o que torna uma virtude virtuosa e um vício vicioso é principalmente uma questão de seus resultados práticos, os responsabilistas, como Zagzebski,¹⁴ Baehr¹⁵ e Tanesini¹⁶ pensam que se trata principalmente de uma questão da psicologia do agente.

Os autores no primeiro grupo (confiabilistas) veem a virtude como sendo uma questão de alcançar bons efeitos ou produzir bons resultados. No entendimento deles, virtudes são virtuosas na medida em que são um bom meio para se atingir resultados relevantes. A taxa de sucesso não precisa ser absoluta, mas deve ser razoável. Segue-se desta visão, para usar os exemplos de Battaly,¹⁷ que um agiota que faz doações substanciais recor-

⁸ CASSAM, Quassim. Vice epistemology, in *The Monist*, v. 99, 2016, p. 159-180.

⁹ CASSAM, Quassim. Epistemic insouciance, in *Journal of Philosophical Research*, v. 43, 2018, p. 1-20

¹⁰ BATTALY, Heather. Testimonial Injustice, Epistemic Vice, and Vice Epistemology. In: I. Kidd, G. Polhaus, e J. Medina (Orgs.). *The Routledge Handbook of Epistemic Injustice*. New York: Routledge, 2017, p. 223-231

¹¹ SWANK, Casey. Epistemic Vice. In: G. Axtell (Org.) *Knowledge, Belief, and Character: Readings in Virtue Epistemology*. Lanham: Rowman and Littlefield Publishers, 2000, p. 195-204.

¹² CASSAM, *op. cit.*

¹³ CAMPOS, Veronica. *Akrasia Epistêmica e Ineficiência Epistêmica: uma abordagem baseada em virtudes*. Belo Horizonte: PPGFil UFMG, 2024.

¹⁴ ZAGZEBSKI, *Virtues of the Mind*.

¹⁵ BAEHR, *The Inquiring Mind*.

¹⁶ TENESSINI, Alessandra. Epistemic Vice and Motivation, in *Metaphilosophy*, v. 49(3), 2018, p. 350-367.

¹⁷ BATTALY, Heather. Varieties of Epistemic Vice. In: J. Matheson e R. Vitz (Orgs.). *The Ethics of Belief*. Oxford: Oxford University Press, 2014, p. 52-53.

rentes a instituições de caridade tem a virtude da benevolência, mesmo que seja motivado apenas pela preocupação egoísta de obter descontos no imposto de renda do ano seguinte. Da mesma forma, um aluno que adquire uma porção de crenças verdadeiras através da leitura repetitiva de materiais acadêmicos no dia anterior à prova tem virtudes epistêmicas, mesmo que a sua única motivação seja a aprovação no exame final, em vez de efetivamente aprender.

Os autores no segundo grupo, por outro lado (os motivacionalistas), entendem que meramente obter bons resultados não torna uma pessoa excelente e não é, portanto, índice de virtude. Para eles, o porquê de alguém ter alcançado ou estar tentando alcançar esses resultados também importa, ou seja, é preciso ter os motivos adequados. Uma taxa de sucesso notável na obtenção de bons resultados não é suficiente e pode até não ser necessária. Portanto, nem o agiota do exemplo acima possui a virtude da benevolência, nem o aluno que se preocupa apenas em ser aprovado no exame final (mas não com aprender) possui qualquer virtude epistêmica. Os motivacionistas discordam entre si sobre se ter os motivos adequados é suficiente para a virtude, mas de modo geral concordam que é necessário, e uma dinâmica análoga vale para os vícios epistêmicos, nos dois casos: para os confiabilistas, o potencial de provocar efeitos epistêmicos negativos torna uma característica um vício, ao passo que para os responsabilistas isso é uma questão da presença de motivações inadequadas, isto é, quando o agente não está preocupado com o conhecimento (em contextos em que deveria estar), mas sim com outras coisas menos importantes.

Munidos dessa compreensão instrumental preliminar sobre o que são virtudes e vícios epistêmicos, e sobre como opera a epistemologia baseada em virtudes, voltemos nossa atenção ao ChatGPT. Importa ganhar clareza sobre quais efeitos o uso do ChatGPT está produzindo em seus usuários, bem como sobre se o uso dessa ferramenta tem alterado para melhor ou para pior as motivações deles, uma vez que os elementos centrais à configuração de uma característica pessoal como uma virtude ou como um vício são, como dissemos, a obtenção de resultados positivos ou negativos (nocivos), e a posse de motivações nobres ou torpes.

2. Bônus: possíveis ganhos epistêmicos ao usar o ChatGPT

A pesquisa científica e a educação são duas práticas humanas que, por excelência, visam à aquisição e à transmissão do saber, respectivamente; por isso examinaremos, a seguir, o uso do ChatGPT em contextos científicos e pedagógicos.

O ChatGPT rapidamente atingiu ampla aceitação dentro da comunidade científica. De acordo com Vilas Boas & Vilas Boas,¹⁸ desde o final de 2022, revistas científicas de renome como *Nature* já reconheciam o uso do ChatGPT como assistente de pesquisa, em tarefas como organização de ideias e resumos da literatura, desde que cumprida uma exigência de transparência (os autores devem esclarecer se um LLM foi utilizado e em que etapa do processo). Apesar de o uso de LLMs como o ChatGPT ser aceito, tais tecnologias não podem ser listadas como autoras. De acordo com os critérios do *International Committee of Medical Journal Editors* (ICMJE), ser coautor demanda, entre outros requisitos, aprovação final da versão publicada e responsabilidade pelo conteúdo, algo que uma IA não pode assumir.

Vilas Boas & Vilas Boas¹⁹ reconhecem que essas ferramentas podem transformar positivamente a prática científica, destacando benefícios como agilidade, inclusão e estímulo à inovação. O ChatGPT permite ao cientista trabalhar mais rápido: ao automatizar tarefas repetitivas como escrita de rascunhos, revisão de literatura ou estruturação inicial de textos, o ChatGPT pode acelerar o ciclo de produção científica. Ele também auxilia na inclusão, ao facilitar a redação para quem não tem a língua inglesa como língua materna; e fomenta a inovação, na medida em que libera tempo e energia dos pesquisadores, de modo que sobra mais espaço para criatividade, intercâmbio e novas ideias.

Possivelmente em razão de vantagens como essas, o ChatGPT está mudando a maneira como os cientistas trabalham. Um estudo de Kobak e colaboradores²⁰ examinou 14 milhões de artigos científicos publicados entre 2010 e 2024 e encontrou uma mudança repentina e drástica que ocorreu logo após o surgimento do ChatGPT, em 2022. A maneira como eles fizeram essa descoberta foi medindo picos na ocorrência de certas palavras, em relação a episódios importantes na ciência. Por exemplo, a frequência da palavra “Ebola” atingiu o pico em 2015, após o surto de Ebola, enquanto a frequência da palavra “zika” atingiu o pico em 2017, após a epidemia do vírus zika. Em 2020, uma das maiores mudanças aconteceu, com palavras como “lockdown”, “pandemic” e “respiratory” atingindo o pico durante o surto de covid, no que é reconhecido como um dos maiores impactos na publicação científica da história. Agora, após 2022, as palavras que apresentam o maior aumento são palavras como “delves”, “crucial” e “potential”. Curiosamente, essas não são palavras relacionadas ao conteúdo científico de um artigo, mas sim ao seu estilo de escrita. São exatamente o tipo de palavras favorecidas pelos LLMs. O aumento sem precedentes

¹⁸ VILAS BOAS, Paulo José; VILAS BOAS, José Vitor. The use of ChatGPT in scientific publishing. *Geriatric, Gerontology and Aging*, 2023, 17:e0230027.

¹⁹ VILAS BOAS & VILAS BOAS, The use of ChatGPT in scientific publishing, p. 3-4.

²⁰ KOBAK, Dmitry; MÁRQUEZ, Rita González; HORVÁT, Eموke-Agnes; LAUSE, Jan. Delving into ChatGPT usage in academic writing through excess vocabulary, in *Science Advances*, v. 11, n. 27, 2025.

de palavras com excesso de estilo após o lançamento do ChatGPT e de outros LLMs nos permite usá-las como marcadores do uso do ChatGPT. Os autores entendem que isso significa que os LLMs estão mudando o discurso científico em uma escala sem precedentes na história.

Numa elaboração distintamente filosófica, Alvarado²¹ propôs recentemente que LLMs devem ser considerados como *tecnologias epistêmicas*. Esse conceito merece explicação: tecnologias epistêmicas são aquelas projetadas especificamente para melhorar nossas capacidades como conhecedores, ou seja, são ferramentas destinadas a auxiliar na obtenção do conhecimento. Segundo o autor, muitas tecnologias – digamos, um martelo ou um smartphone – não são primariamente destinadas à produção de conhecimento. Já as tecnologias epistêmicas têm a investigação, análise e manipulação de dados (tarefas distintamente epistêmicas) como suas tarefas centrais. Assim, o ChatGPT é uma tecnologia epistêmica porque possui características destacadas: ele é projetado para ser empregado em atividades de investigação, busca ou inquisição, isto é, atividades que tem como objetivo gerar ou refinar conhecimento; ele manipula conteúdo epistêmico, isto é, lida diretamente com dados, informação e representações (entidades que são, por definição, elementos do conhecimento); faz isso de forma mais rápida do que um agente epistêmico humano faria; e seu *modus operandi* é distintamente epistêmico, já que suas principais funções são previsão e análise, isto é, reconhecimento de padrões, inferência, explicação de dados, dentre outras. Essas operações correspondem diretamente ao trabalho epistêmico de formular hipóteses, detectar padrões e extrair conclusões.

Se LLMs são uma tecnologia epistêmica no sentido delineado acima, eles são uma ferramenta para investigadores, o que inclui cientistas, na medida em que absorvem parte do trabalho epistêmico que de outro modo precisaria ser feito “à mão”. Mas é difícil determinar se esse impacto de um ponto de vista qualitativo, quer dizer, é difícil especificar se as mudanças no fluxo do trabalho epistêmico de investigadores e cientistas são predominantemente positivas, ou até que ponto elas o são. Na área da educação, a realidade não é tão diferente: o ChatGPT oferece uma série de benefícios a seus usuários, e passou a ser adotado praticamente desde seu lançamento, de modo que ele vem se tornando uma espécie de “tecnologia pedagógica”, em linhas semelhantes àquelas por meio das quais ele foi descrito como uma tecnologia epistêmica.

Autores como Dawson e colaboradores²² apontaram que uma das dificuldades de se examinar o impacto de ferramentas como o ChatGPT e outras

²¹ ALVARADO, Ramón. AI as an Epistemic Technology, in *Science and Engineering Ethics*, 29(32), 2023.

²² DAWSON, Shane; JOKSIMOVIC, Srecko; MILLS, Caitlin; GASEVIC, Dragan; SIEMENS, George. Advancing theory in the age of artificial intelligence, in *British Journal of Educational Technology*, v. 54, 2023, p. 1051-1056.

similares sobre a transmissão de conhecimento é dada pelo fato de que não temos **bases teóricas** sensíveis à coevolução entre essas tecnologias e experiências educativas. Como os autores enfatizam, a maioria das teorias educacionais emergiu em contextos anteriores à atual onda de IA generativa. Estamos usando a IA em contextos pedagógicos, mas sem que esse uso seja balizado por uma compreensão robusta do papel que essa tecnologia desempenha (ou deveria desempenhar) na educação. Ou seja, estamos incorporando uma ferramenta à educação antes de termos grande clareza sobre por quê ela deveria estar ali e, obviamente, sem grande clareza sobre a proporção dos benefícios e dos riscos.

Os benefícios oferecidos por uma tecnologia podem ser entendidos em termos de *affordances*. Esse é um termo oriundo do repertório conceitual da psicologia ecológica, que basicamente se refere ao leque de possibilidades que o acesso a um determinado recurso “abre” ao indivíduo. Ou seja, trata-se de ganhos potenciais que o recurso é capaz de viabilizar. Diversos trabalhos recentes têm, de forma consideravelmente otimista, chamado a atenção para as *affordances* relacionadas à incorporação da IA generativa ao processo de ensino-aprendizagem em diferentes áreas, entre os quais Eager & Brunton,²³ Järvelä, Nguyen & Hadwin,²⁴ Sharples,²⁵ Cain,²⁶ Fuertes-Alpiste,²⁷ Mishra, Oster & Henriksen,²⁸ Simms,²⁹ entre outros.

Uma das primeiras revisões sistemáticas que examinam o uso educacional emergente do ChatGPT e que destacam particularmente os benefícios associados a ele é o trabalho de Crompton & Burke.³⁰ No que tange às *affordances* ligadas à transmissão do conhecimento (i.e., apoio ao desempenho docente) do ChatGPT, esse estudo destaca três categorias principais: suporte ao ensino, automação de tarefas e desenvolvimento profissional. O ChatGPT emergiu como ferramenta capaz de auxiliar na elaboração de conteúdos, planejamentos de aulas e atividades, personalizar experiências

²³ EAGER, Bronwyn; BRUNTON, Ryan. Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice, in *Journal of University Teaching & Learning Practice*, 20(5), 2023.

²⁴ JARVELA, Sanna; NGUYEN, Andy; HADWIN, Allyson. Human and artificial intelligence collaboration for socially shared regulation in learning, in *British Journal of Educational Technology*, 54, 2023, p. 1057–1076.

²⁵ SHARPLES, Mike. Towards social generative AI for education: theory, practices and ethics, in *Learning: Research and Practice*, 9(2), 2023, p. 159–167.

²⁶ CAIN, William. Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education, in *TechTrends*, 68, 2024, p. 47–57.

²⁷ FUERTES-ALPISTE, Marc. Enmarcando las aplicaciones de IA generativa como herramientas para la cognición en educación Pixel-Bit, in *Revista De Medios Y Educación*, n. 71, 2024, p. 42–57.

²⁸ MISHRA, Punya; OSTER, Nicole; HENRIKSEN, Danah. Generative AI, Teacher Knowledge and Educational Research: Bridging Short- and Long-Term Perspectives, in *TechTrends* n. 68, 2024, p. 205–210.

²⁹ SIMMS, Rachells. Work With ChatGPT, Not Against: 3 Teaching Strategies That Harness the Power of Artificial Intelligence, in *Nurse Educator*, 49(3), 2024, p. 158–161.

³⁰ CROMPTON, Helen; BURKE, Diane. The Educational Affordances and Challenges of ChatGPT: State of the Field, in *TechTrends* n. 68, 2024, p. 380–392.

de aprendizado, adaptar materiais ao nível dos alunos e servir de plataforma de reflexão pedagógica para professores. Esses usos ilustram uma versatilidade que pode aliviar parte da carga administrativa e criativa do docente, permitindo-lhe concentrar-se mais no processo educativo propriamente dito, dedicando mais tempo ao ensino reflexivo e humanizado.

Ainda de acordo com a revisão conduzida por Crompton & Burke,³¹ do ponto de vista do aluno, os usos identificados também são diversificados. O ChatGPT pode oferecer suporte ininterrupto (24/7), explicar conceitos complexos de forma acessível, atuar como interlocutor em debates textuais, fornecer feedback personalizado e materiais adaptados, auxiliar na produção textual, apoiar o autoestudo e a autoavaliação, fomentar o engajamento com o conteúdo e incentivar a autodeterminação. Essa gama de funcionalidades evidencia que a ferramenta pode desempenhar múltiplos papéis no ecossistema educacional, adaptando-se tanto às necessidades de scaffolding cognitivo quanto a aspectos motivacionais do processo de aprendizagem.

Mais recentemente, Rivera-Novoa & Duarte Arias³² argumentaram que, em determinados contextos, tecnologias como o ChatGPT e congêneres podem funcionar como um **artefato cognitivo complementar**, promovendo o aprendizado ativo, partindo do framework oferecido pelas teorias cognitivas da mente estendida (por exemplo, Clark & Chalmers³³). Essas teorias entendem artefatos como “extensões” da cognição humana – um bloco de anotações, por exemplo, pode funcionar como uma “extensão” da memória de uma pessoa com Alzheimer; assim como um calendário digital com funções de lembretes pode funcionar como uma extensão de minhas faculdades de planejamento. O trabalho expõe três casos com suporte empírico: (i) a produção de feedback automatizado; (ii) tecnologias assistivas para estudantes com necessidades especiais; e (iii) aplicações gamificadas que utilizam IA generativa para produzir interações personalizadas.

Segundo os autores, no primeiro caso, ferramentas que fornecem feedback formativo, como modelos multiagentes ou chatbots socráticos, permitem que os aprendizes identifiquem lacunas em sua compreensão e reflitam criticamente sobre suas respostas. Esses sistemas amplificam e orientam o processo cognitivo, num arranjo cognitivo estendido em que o aprendiz participa de forma ativa. Já o segundo caso concentra-se em tecnologias assistivas: o ChatGPT pode ser empregado para produzir imagens acessíveis para pessoas com aphantasia (dificuldade para imaginar cenas),

³¹ CROMPTON & BURKE, The Educational Affordances and Challenges of ChatGPT: State of the Field.

³² RIVERA-NOVOA, Angel; DUARTE ARIAS, Daniel. Generative Artificial Intelligence and Extended Cognition in Science Learning Contexts, in *Science & Education* 2025.

³³ CLARK, Andy; CHALMERS, David. The Extended Mind, in *Analysis*, v.58, n. 1, 1998, p. 7–19.

transformações em texto para estudantes com deficiências motoras, ou descrições de imagens para deficientes visuais. De acordo com Rivera-Novoa & Duarte Arias,³⁴ essas aplicações não substituem o processo mental, ao contrário, elas oferecem suporte para que o aluno continue exercendo as próprias capacidades cognitivas, integrando o aparelho tecnológico como parte do seu sistema estendido de cognição. O terceiro exemplo evidencia a eficácia de abordagens gamificadas. Ferramentas como jogos personalizados para estudantes com dislexia ou ambientes multiagentes de debate de perspectivas favorecem interações dinâmicas, motivação e engajamento crítico. O sistema gera conteúdo adaptado, mas requer que o aluno participe ativamente, tornando-se, assim, parte integrada de um circuito cognitivo estendido e harmônico. Esses casos ilustram variedades de uso discutivelmente virtuoso do ChatGPT em contextos distintamente pedagógicos na esperança de que, assim, o aluno possa aprender mais e melhor.

3. Ônus: possíveis custos epistêmicos, individuais e coletivos

Mas nem tudo sobre o ChatGPT são flores. Dentre os problemas recorrentes observados nos estudos revisados por Crompton & Burke,³⁵ destacam-se principalmente imprecisões e as “alucinações” (fenômenos em que o ChatGPT gera afirmações incorretas, imprecisas ou sem respaldo factual e as apresenta como se fossem corretas, precisas e factuais), e restrições operacionais próprias da ferramenta, quer dizer, limites em termos de execução, ou executabilidade, de tarefas. Além desses problemas, há também os vieses embutidos na base de dados com que essas ferramentas são treinadas e há a questão sobre que tipo de impacto sobre o usuário (o que inclui impacto sobre sua saúde e impacto cognitivo) a ferramenta está provocando. Iremos nos concentrar nesses últimos dois itens.

Em relação ao problema dos vieses, vários estudos vêm sendo feitos. Um estudo recente feito por Bai e colaboradores³⁶ revelou que mesmo LLMs programados para estarem alinhados com valores humanos exibem vieses estereotipados implícitos em categorias como raça, gênero, religião e

³⁴ RIVERA-NOVOA & DUARTE ARIAS, *Generative Artificial Intelligence and Extended Cognition in Science Learning Contexts*.

³⁵ CROMPTON & BURKE, *The Educational Affordances and Challenges of ChatGPT: State of the Field*.

³⁶ BAI, Xuechunzi; WANG, Angelina; SUCHOLUTSKY, Ilia; GRIFFITHS, Thomas. Explicitly unbiased large language models still form biased associations. In *Proceedings of the National Academy of Sciences of the United States of America*, 122 (8), 2025, e2416228122.

saúde. Numa linha semelhante, experimentos com cadeias de transmissão (*transmission chains*) feitos por Acerbi & Stubbersfield³⁷ mostraram que o ChatGPT tende a preservar conteúdos que são consistentes com este-reótipos de gênero, além de preferir informações sociais, negativas ou relacionadas a ameaça, evidenciando vieses cognitivos comuns. Outros trabalhos (por exemplo, Rickman;³⁸ Kirk e colaboradores³⁹) revelaram que modelos como ChatGPT e Bard continuam a propagar informações médicas racistas e sexistas. Juntos, os estudos que têm encontrado esse tipo de problema com os LLMs sugerem a necessidade de considerarmos *o que* esses modelos deveriam estar aprendendo – se eles devem ajudar a corrigir ou se devem apenas reproduzir desigualdades e injustiças já presentes na sociedade. Um custo cognitivo óbvio de permitir que eles simplesmente reproduzam vieses é que, à medida que as pessoas recorrem ao ChatGPT para se informar, tais vieses serão reforçados. Ou seja, via mecanismos de reforço da mentalidade preexistente, nós, como indivíduos e como sociedade, tenderemos a nos tornar ainda mais preconceituosos e discriminatórios do que já somos, no médio e longo prazos – sendo esses traços vícios epistêmicos por definição.

Sem dúvida é preocupante que sistemas de inteligência artificial possam reproduzir práticas racistas, sexistas, etaristas, entre outras, mas é importante não perder de vista que tais condutas não são exclusivas dessas tecnologias, nem foram introduzidas em nossa cultura com a introdução delas. Essa são práticas que os próprios seres humanos realizam cotidianamente. Também é importante ter em mente que sistemas de inteligência artificial não possuem um grau de sofisticação tão elevado quanto, por vezes, se sugere. Embora alguns artigos enfatizem o caráter “assustador” de um bot poder manifestar traços racistas ou imorais, tais comportamentos derivam, em grande medida, da tentativa de imitar o funcionamento humano. O que deveria ser primeiramente assustador é que tenha havido, ou que haja, *pessoas* com esses traços de caráter (ou vícios) tão detestáveis, que a inteligência artificial esteja emulando.

Em relação ao problema do impacto cognitivo, ainda há poucos estudos investigando esses aspectos, mas um deles é um estudo recente do MIT⁴⁰

³⁷ ACERBI, Alberto; STUBBERSFIELD, Joseph. Large language models show human-like content biases in transmission chain experiments. In *Proceedings of the National Academy of Sciences of the United States of America*, 120(44), 2023, e2313790120.

³⁸ RICKMAN, Sam. Evaluating gender bias in large language models in long-term care, in *BMC Medical Informatics and Decision Making*, 25(274), 2025.

³⁹ KIRK, Hannah; JUN, Yennie; IQBAL, Haider; BENUSSI, Elias; VOLPIN, Filippo; DREYER, Frederic; SHTEDRITSKI, Aleksandar; ASANO, Yuki. Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. 2021, *35th Conference on Neural Information Processing Systems*.

⁴⁰ KOSMYNA, Nataliya; HAUPTMAN, Eugene; YUAN, Ye Tong; SITU, Jessica; LIAO, Xian-Hao; BERESNITZKY, Ashly; BRAUNSTEIN, Iris; MAES, Pattie. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 2025. Disponível em: arXiv preprint arXiv:2506.08872

que avaliou o uso do ChatGPT para escrever ensaios e encontrou alguns resultados preocupantes. O estudo gerou uma série de manchetes sensacionalistas, como a reportagem “O ChatGPT pode estar corroendo as habilidades de pensamento crítico”, publicada pela Time Magazine⁴¹ em junho de 2025, e “Pesquisadores do MIT dizem que usar o ChatGPT pode ‘apodrecer’ seu cérebro”, publicada pelo G1 em julho do mesmo ano.⁴² Sensacionalismo à parte, as descobertas feitas pelo estudo foram menos sobre deterioração cerebral, num sentido literal, e mais sobre o uso de atalhos mentais e suas possíveis consequências deletérias.

Nesse estudo, os pesquisadores dividiram 54 estudantes universitários da região de Boston em três grupos e deram a cada um dos grupos uma tarefa de redação. O primeiro grupo usou o ChatGPT para executar a tarefa, o segundo grupo usou o Google (sem a ajuda da IA) e o terceiro teve que executar a tarefa sem usar tecnologias digitais. Enquanto escreviam, a atividade cerebral foi medida por eletrodos. Após três rodadas, o grupo que não teve acesso à internet apresentou os maiores níveis de conectividade mental, ao passo que o grupo de usuários do ChatGPT apresentou os níveis mais baixos. Na quarta rodada, os papéis do primeiro e do terceiro grupos se inverteram. O grupo que havia executado as tarefas de redação sem usar tecnologias pôde usar o ChatGPT, enquanto o grupo que anteriormente havia usado o ChatGPT teve que trabalhar sem usar nenhuma tecnologia. O resultado foi que os participantes do grupo que havia se “acostumado” a ter que escrever sem usar tecnologia alguma melhoraram a qualidade de suas redações, ao passo que aqueles que haviam se “acostumado” a usar o ChatGPT tiveram grande dificuldade até mesmo para se lembrar do que haviam escrito.

Uma forma de interpretar os achados desse estudo é entender que usuários que recorrem ao ChatGPT como primeira medida para realizar tarefas intelectuais tendem a ter pior desempenho cognitivo do que aqueles que não usam a ferramenta, ou usam-na não como primeiro recurso, mas como recurso adicional para aprimorar tarefas que foram feitas “analogicamente”. Entre os principais efeitos de piora cognitiva, um dos mais alarmantes é a aparente capacidade do ChatGPT de piorar a capacidade de memorização do usuário. Essa interpretação é compatível com a observação feita por professores de inglês que revisaram os trabalhos entregues pelos participantes de cada grupo, e disseram que os textos gerados pelo ChatGPT eram desprovidos de originalidade, individualidade e de “espírito” (*soul*):

Algumas redações, abrangendo todos os temas, se destacaram pelo uso quase perfeito da linguagem e da estrutura, enquanto, ao mesmo tempo, falharam em apresentar *insights* pessoais ou declarações claras. Essas re-

⁴¹ Disponível em <https://shorturl.at/0LTaq>

⁴² Disponível em <https://shorturl.at/b6VjS>

dações, muitas vezes longas, incluíam ideias padronizadas, formulações e declarações típicas recorrentes, o que tornou o uso de IA no processo de escrita bastante óbvio. Nós, como professores de inglês, percebemos essas redações como “sem espírito”.⁴³

Se o indivíduo não precisar pensar para executar uma tarefa na qual ele não é proficiente, é compreensível que o resultado deixe transparecer isso. No geral, ao longo dos quatro meses em que o estudo foi realizado, os participantes que utilizaram o ChatGPT desde o início ficaram aquém dos demais grupos em termos de níveis neural, linguístico e comportamental, além de terem dedicado menos tempo às suas redações, simplesmente clicando em “copiar” e “colar”. Isso foi conceitualizado pelos autores em termos de “débito cognitivo”. Trata-se do acúmulo de uma espécie de “dívida”, que foi definida nos seguintes termos:

[débito cognitivo é o] uma condição na qual a dependência repetida de sistemas externos, como os LLMs, substitui os processos cognitivos trabalhosos necessários para o pensamento independente. A dívida cognitiva adia o esforço mental a curto prazo, mas resulta em custos a longo prazo, como diminuição da investigação crítica, maior vulnerabilidade à manipulação e diminuição da criatividade. Quando os participantes reproduzem sugestões sem avaliar sua precisão ou relevância, não apenas perdem a propriedade das ideias, mas também correm o risco de internalizar perspectivas superficiais ou tendenciosas.⁴⁴

Na seção “discussão” do artigo, Kosmyrna e colaboradores *relacionam o débito cognitivo* com implicações mais amplas, especialmente no campo educacional. Um dos primeiros pontos enfatizados é a diferença de intensidade no envolvimento cognitivo observada entre os grupos estudados. Os participantes que escreveram seus textos sem qualquer apoio de ferramentas digitais apresentaram padrões de atividade cerebral mais intensos e distribuídos, em particular nas bandas de frequência associadas à concentração e ao esforço mental. Esses resultados confirmam a hipótese intuitiva de que o ato de escrever, por si (sem apoio de tecnologias) demanda maior mobilização cognitiva do que o ato de delegar a redação ao ChatGPT. Acresça-se a isso a observação – particularmente destacada pelos autores – de que a diminuição do engajamento cognitivo não desaparece imediatamente quando os participantes deixam de usar o ChatGPT, e por isso ela tende a se acumular, criando uma “dívida”. A dívida é todo o exercício que o agente precisaria ter feito para ser bem-sucedido na realização de uma tarefa mas que, graças ao ChatGPT, não precisou fazer. É como deixar de ir à academia, dia após dia – com a iteração, acumula-se

⁴³ KOSMYRNA et. al., Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, p. 62.

⁴⁴ KOSMYRNA et. al., Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, p. 141.

um “débito”, no sentido de que é como se o indivíduo estivesse “devendo” todos os exercícios físicos que não fez.

O estudo mostrou que aqueles que passaram de uma condição de escrita com auxílio do ChatGPT para uma condição sem suporte digital continuaram a apresentar níveis reduzidos de conectividade nas frequências alfa e beta, mesmo após terem abandonado a ferramenta. Esse resultado sugere que o impacto do uso da inteligência artificial na atividade cerebral pode ser mais duradouro do que se poderia imaginar, de modo que a “dívida cognitiva” não se desfaz de maneira instantânea. Tal persistência é interpretada como um indício de infraengajamento continuado, que compromete a plena retomada da atividade cognitiva característica da escrita independente. Isso significa que, com a repetição (e a médio e longo prazo), pode-se esperar que capacidades cognitivas ligadas à escrita, por uma questão de desuso, tendam a “atrofiar” nos indivíduos que delegam tarefas ao ChatGPT como primeiro recurso.

Outro achado relevante discutido no artigo refere-se à relação entre uso de modelos de linguagem e a experiência subjetiva de autoria. Os participantes que utilizaram o ChatGPT para compor seus ensaios relataram um senso de autoria significativamente menor em comparação com aqueles que escreveram sem suporte. Além disso, esses indivíduos demonstraram maior dificuldade em recordar ou citar corretamente trechos de seus próprios textos. Esse fenômeno sugere que a delegação parcial ou integral da escrita a um agente artificial não apenas reduz o esforço cognitivo no momento da produção, mas também compromete a consolidação da memória e a identificação pessoal com o resultado final. Em termos educacionais, esse enfraquecimento da autoria pode significar não apenas uma diminuição do aprendizado profundo, mas também a erosão da responsabilidade intelectual do indivíduo em relação ao que produz.

Por fim, a seção de discussão aborda certos aspectos sociais mais amplos desses achados. Embora os modelos de linguagem ofereçam benefícios inegáveis em termos de rapidez, praticidade e capacidade de apoiar a produção textual, o uso prolongado ou excessivo pode diminuir a aprendizagem profunda, enfraquecer da memória e criar uma cultura de menor disposição ao esforço intelectual necessário para tarefas complexas. Os autores enfatizam que, em contextos educacionais, a dependência acrítica desses sistemas pode reduzir o engajamento cognitivo de toda uma geração, o que comprometeria a qualidade do aprendizado em longo prazo e resultaria em indivíduos adultos possivelmente mais preguiçosos, ingênuos e menos cognitivamente capazes. Essa advertência é especialmente pertinente em um momento histórico no qual a integração de sistemas de inteligência artificial em ambientes de ensino está se acelerando rapidamente, muitas vezes sem avaliação criteriosa dos efeitos cognitivos.

Pensando a partir de um framework de epistemologia baseada em virtudes, o que podemos retirar disso é que o uso indiscriminado do ChatGPT tem um potencial para gerar prejuízos epistêmicos, na medida em que nos prejudica em nossas capacidades como conhecedores. Se as coisas persistirem como estão, devemos ter que lidar, num futuro próximo, com uma geração de pessoas intelectualmente mais passivas: menos interessadas e dispostas a investigar, menos capazes de investigar bem, além de mais acríticas e propensas a apresentar como de sua autoria trabalhos não foram efetivamente feitos por elas. Esses traços podem ser conceitualizados em termos de vícios como a preguiça, a indiferença intelectual (*epistemic insouciance*), a ingenuidade e a desonestidade. Se acrescentarmos a isso a constatação de que o ChatGPT reproduz vieses, teremos também pessoas mais preconceituosas. O desafio que se coloca, nesse sentido, é encontrar formas de integrar a inteligência artificial de modo que ela complemente, mas não substitua, o esforço cognitivo humano. Isso implica pensar em estratégias pedagógicas que incentivem o estudante a manter o protagonismo intelectual, utilizando a IA como apoio, mas não como substituto integral da atividade de reflexão, memória e autoria. Esse ponto nos leva diretamente ao tópico seguinte, sobre distorção de propósitos.

4. Distorção de propósitos

Ao delegar tarefas epistêmicas, como a tarefa de escrita, para o ChatGPT, fazemos uma troca. Estamos trocando o benefício do processo pelo benefício do resultado. Em muitas ocasiões, essa troca pode valer a pena, desde que conheçamos o processo e já possuamos a proficiência nas habilidades que de outro modo o processo de fazer sozinho iria fortalecer. Para fazer uma analogia simples: não estamos perdendo muita coisa ao delegar a função de lavar a louças para uma máquina lava-louças – mas isso é porque todos nós (eu espero) somos proficientes na habilidade de lavar louça. As competências do processo já estão suficientemente desenvolvidas, de modo que não se faz rigorosamente vantajoso continuar a exercitá-las para fins de aprimoramento. Nesse caso, podemos abrir mão dos benefícios do exercício em prol do benefício do resultado (a máquina entregará a louça limpa, dentro de algumas horas, e nesse período talvez possamos nos dedicar a coisas mais importantes, como estudar uma língua, entrar em contato com a arte e socializar). Estamos sacrificando o processo de lavar a louça e pulando direto para o resultado de ter a louça limpa dentro de algumas horas, e esse sacrifício é mais vantajoso do que desvantajoso para nós.

Por que, alguém poderia perguntar, o mesmo não se dá no caso da escrita? Por que não se trata simplesmente de uma questão de prioridades? Uma pessoa pode priorizar aprender uma língua estrangeira ou entrar em

contato com alguma forma de arte, ou mesmo socializar, em detrimento de lavar a louça manualmente, e isso é visto como perfeitamente positivo. Por que não pode uma pessoa priorizar fazer qualquer outra atividade em detrimento de fazer tarefas de escrita manualmente? Qual exatamente é o problema, que torna a delegação aceitável no primeiro caso e inaceitável (conforme estamos argumentando) no segundo?

O aspecto mais relevante a ser ressaltado quanto a isso é a facilidade com que essas ferramentas podem comprometer, piorar, diminuir ou mesmo estragar nossa qualidade enquanto agentes epistêmicos. Uma trajetória epistêmica é uma trajetória que nos leva da ignorância ao conhecimento, ou entendimento. Ou seja, é uma jornada por meio da qual mudamos de um ponto A, em que não sabemos ou não compreendemos uma coisa, até um ponto B, em que descobrimos ou temos uma compreensão razoável aquilo que pretendíamos compreender. Nós somos o resultado de todas as jornadas epistêmicas que já trilhamos, isto é, nós melhoramos epistemicamente na medida em que trilhamos nossas jornadas em busca de conhecimento. Nós adquirimos habilidades e competências ao trilhar a jornada, e exercitamos essas habilidades e competências, nos tornando investigadores melhores e mais aptos a trilhar jornadas mais complexas e mais longas no futuro.

Um ponto sobre jornadas epistêmicas é que elas sempre são constituídas por uma série de tarefas, sendo algumas mais simples e outras mais complexas, isto é, cognitivamente demandantes. O advento do ChatGPT torna tentador transferir para a máquina as tarefas cognitivamente mais exigentes da jornada epistêmica, dada a acessibilidade, a rapidez e a precisão oferecidas. Essa atitude de delegar as tarefas mais demandantes vai aprofundando um padrão de dependência, uma dependência que tende a dificultar a autonomia intelectual. Muitos jovens, por exemplo, não exercitam mais cálculos no papel e não são mais capazes de fazê-lo, recorrendo imediatamente ao celular. Do mesmo modo, há crianças que interagem quase exclusivamente com dispositivos digitais, deixando de cultivar a imaginação em brincadeiras concretas ou mesmo de manter contato ativo com a leitura tradicional de livros. O problema a ser observado aqui não é que o uso da calculadora do celular seja nocivo *em si* mesmo, mas sim que o hábito continuado de depender desse dispositivo nos leva numa espécie de “jornada epistêmica invertida” – como resultado dos esforços poupados, deixamos de exercitar habilidades e competências e, conseqüentemente, as perdemos. Nós pioramos epistemicamente.

Assim como calcular, escrever é uma habilidade extremamente importante que os humanos devem aprender e aprimorar, a fim de melhorarem ou, pelo menos, manterem sua qualidade como agentes epistêmicos. Um primeiro motivo para isso é o valor intrínseco da escrita. Todos deveriam

saber se comunicar por ser essa uma das coisas mais fundamentais que fazemos, e uma das mais criativas. Ou seja, é uma das atividades que nos tornam humanos. Mas um segundo motivo, e um que é mais relevante para essa discussão, é o valor instrumental da escrita. Escrever serve para organizar o pensamento, como diversos estudos têm mostrado. Por exemplo, Kellogg⁴⁵ mostrou que ato de escrever amadurece desde a “contagem de conhecimentos” até o “trabalho/refinamento do conhecimento”, operando em múltiplos níveis cognitivos simultaneamente; enquanto Hillesund⁴⁶ mostrou que leitura e escrita estão intrinsecamente interligadas como forma de cognição distribuída, de modo que escrever molda novas ideias ao longo do processo e faz com que ganhemos clareza sobre as ideias que já tínhamos. No caso da escrita, portanto, abrir mão do processo em favor do resultado configura uma distorção de propósitos. Aquilo de que estamos abrindo mão não é só o esforço envolvido no processo – é também um rol de benefícios epistêmicos que esse esforço nos daria. Ao colocar o ChatGPT para pensar por nós estamos escolhendo, em outras palavras, nos tornar pessoas não tão boas em nossa capacidade de pensar por nós mesmos. Ou, em termos um pouco mais dramáticos: ao colocar uma inteligência artificial para trabalhar em nosso lugar em tarefas epistêmicas, estamos escolhendo nos tornar, nós próprios, menos inteligentes.

Um outro ponto relacionado tem a ver com a pergunta sobre quem está se beneficiando dessa delegação em massa de tarefas epistêmicas para a inteligência artificial. Quando utilizamos o ChatGPT para contornar nossa incapacidade de escrever textos decentes, quem se beneficia com isso, de verdade? Não somos nós já que, como venho argumentando, os prejuízos que tomamos ao fazer isso são comparáveis, se é que não superam, os benefícios de fazê-lo. Também não é o indivíduo que irá avaliar o texto, por exemplo, o professor, ou um parecerista. Parece que tal prática só beneficia realmente ao próprio ChatGPT e àqueles que lucram através dele. Isso vale para todas as tecnologias que prometem servir como substitutos da criatividade. Criatividade é uma virtude humana, algo que devemos gostar de praticar e que não deve ser terceirizado ou negligenciado. Tudo o que o ChatGPT expôs foi a dura realidade de que precisamos reavaliar coletivamente nossa compreensão da escrita e melhorar nossa relação com ela. Se o ChatGPT é capaz de escrever melhor do que nós, isso significa que nossos padrões não são altos o suficiente, e que deveríamos estar tentando melhorá-los, e não abaixá-los ainda mais.

⁴⁵ KELLOGG, Ronald. Training writing skills: A cognitive development perspective, in *Journal of Writing Research* 1(1), 2008.

⁴⁶ HILLESUND, Terje. Scholarly reading (and writing) and the power of impact factors: a study of distributed cognition and intellectual habits, in *Frontiers in Psychology*, 17(14), 2023, 1165700.

Considerações finais

Ao longo deste artigo, procurei analisar o ChatGPT e outros LLMs a partir da lente da epistemologia baseada em virtudes. Essa abordagem revelou que tais tecnologias não podem ser avaliadas apenas em termos instrumentais ou de eficiência, mas devem ser examinadas quanto ao impacto que produzem sobre nosso caráter intelectual, isto é, seu impacto sobre nós enquanto agentes epistêmicos. Virtudes intelectuais como curiosidade, humildade, honestidade e diligência são fundamentais para o florescimento de uma vida cognitiva saudável; inversamente, vícios como preguiça, indiferença epistêmica, arrogância e ingenuidade corroem nossas capacidades de conhecer e compreender. O ponto crucial que emerge da investigação é que, embora os LLMs ofereçam inegáveis benefícios práticos, eles também apresentam um potencial significativo de nos infestar com vícios epistêmicos, sobretudo quando usados de modo indiscriminado.

Um primeiro aspecto destacado foi a tendência dessas ferramentas a incentivar a passividade intelectual. Ao automatizar tarefas de escrita, pesquisa e síntese, o ChatGPT reduz o esforço cognitivo necessário para aprender e compreender, levando a um “débito cognitivo”, isto é, levando à perda progressiva de habilidades como memória, pensamento crítico e autoria quando os indivíduos recorrem sistematicamente à IA como primeira estratégia de resolução de problemas. Sob a ótica da epistemologia baseada em virtudes, isso representa a consolidação de vícios como a preguiça intelectual e a indiferença epistêmica: *em vez de cultivarmos diligência e perseverança na investigação, e o gosto por investigar, delegamos o trabalho ao chat, sacrificando o processo de aprendizado em nome da comodidade.*

Outra fonte relevante de cristalização de vícios epistêmicos está associada ao enfraquecimento do senso de autoria e responsabilidade intelectual. Usuários que confiam excessivamente no ChatGPT para redigir textos ou estruturar argumentos tendem a perder a ligação subjetiva com o conhecimento produzido, reconhecendo cada vez menos aquilo que escrevem como fruto de sua própria agência cognitiva. Isso mina virtudes como honestidade e coragem intelectual, e favorece vícios como a desonestidade e a menoridade (dependência, por oposição à emancipação). Em termos educativos, essa dinâmica pode consolidar uma geração de estudantes habituados a “terceirizar” suas tarefas cognitivas, resultando em indivíduos menos autônomos, menos criativos e menos preparados para enfrentar problemas complexos sem auxílio tecnológico.

Não menos grave é o problema dos vieses reproduzidos pelos LLMs. Pesquisas recentes mostram que, mesmo quando programados para serem “neutros”, modelos como o ChatGPT preservam e até amplificam precon-

ceitos de gênero, raça e classe, além de favorecerem conteúdos negativos e estereotipados. Do ponto de vista da epistemologia da virtude, isso reforça vícios como ingenuidade, arrogância e mesmo a malevolência, pois o usuário que confia cegamente no sistema pode aprofundar e agravar crenças enviesadas latentes. Assim, em vez de fomentar curiosidade responsável e humildade intelectual, o uso tende a solidificar preconceitos já existentes, com sérias implicações sociais e epistêmicas.

Esses riscos se tornam ainda mais evidentes quando consideramos a facilidade com que os LLMs promovem uma distorção de propósitos. Ao substituir o processo de investigação, escrita e reflexão pelo simples consumo de resultados prontos, os usuários sacrificam não apenas o esforço envolvido, mas também os benefícios epistêmicos que esse esforço de outro modo traria. A escrita, como mostrado por estudos recentes, é uma atividade intrinsecamente ligada à organização do pensamento e ao desenvolvimento de novas ideias. Abrir mão desse processo equivale, portanto, a abrir mão de um exercício fundamental para o cultivo de virtudes intelectuais. O resultado é uma espécie de “atrofia epistêmica”: indivíduos tornam-se menos capazes de raciocinar, argumentar e aprender de maneira independente. Assim, de um ponto de vista confiabilista em epistemologia da virtude, poderíamos dizer que os ônus do uso indiscriminado do ChatGPT superam, em muitos casos, seus bônus. Já sob a ótica responsabilista, fica claro que os motivos que nos levam a recorrer à IA muitas vezes são falhos: buscamos rapidez e comodidade em detrimento da busca autêntica por conhecimento. Em ambas as perspectivas, portanto, o uso da tecnologia revela-se mais problemático do que sadio.

Não se trata de negar completamente o valor dos LLMs. Eles podem, em certos contextos cuidadosamente regulados, funcionar como tecnologias epistêmicas úteis, ampliando nossas capacidades de investigação e ensino. Contudo, o quadro geral que emerge desta análise é que tais ferramentas carregam um risco substancial de corroer virtudes intelectuais e fomentar vícios. A responsabilidade recai, então, sobre nós, enquanto sociedade, educadores e pesquisadores, de criar ambientes que limitem o uso acrítico e incentivem práticas reflexivas. É preciso desenvolver estratégias pedagógicas e científicas que mantenham o estudante no centro do processo cognitivo, utilizando a IA como apoio, mas nunca como substituto integral do esforço intelectual.

Em suma, os LLMs representam um dilema epistêmico contemporâneo. Podem ser instrumentos valiosos, mas também podem nos tornar piores agentes epistêmicos. A epistemologia baseada em virtudes nos ajuda a enxergar que, se quisermos que tais ferramentas contribuam para nosso florescimento cognitivo, devemos resistir à tentação da delegação indiscriminada e cultivar, em contrapartida, hábitos intelectuais que preservem

a centralidade da agência humana no processo de conhecer. O futuro da inteligência artificial, sob esse prisma, não será definido apenas por avanços técnicos, mas sobretudo por nossa capacidade de usá-la de modo virtuoso, sem abrir mão do esforço epistêmico que nos constitui como seres racionais.

Referências

ACERBI, Alberto; STUBBERSFIELD, Joseph. Large language models show human-like content biases in transmission chain experiments. In: *Proceedings of the National Academy of Sciences of the United States of America*, 120(44), 2023, e2313790120. doi: 10.1073/pnas.2313790120.

ALIER, Marc, GARCÍA-PENALVO, Francisco; CAMBA, Jorge. Generative Artificial Intelligence in Education: From Deceptive to Disruptive, in *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 2024, p. 5-14. doi: 10.9781/ijimai.2024.02.011.

ALVARADO, Ramón. AI as an Epistemic Technology, in *Science and Engineering Ethics*, 2023, 29(32), doi: <https://doi.org/10.1007/s11948-023-00451-3>.

BAEHR, Jason. Epistemic Malevolence. In: H. Battaly (Org.). *Virtue and Vice, Moral and Epistemic*. Chichester: Wiley-Blackwell, 2010, p. 189-213.

BAEHR, Jason. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford, Oxford University Press, 2012.

BAI, Xuechunzi; WANG, Angelina; SUCHOLUTSKY, Ilia; GRIFFITHS, Thomas. Explicitly unbiased large language models still form biased associations. In: *Proceedings of the National Academy of Sciences of the United States of America*, 122 (8), 2025, e2416228122, <https://doi.org/10.1073/pnas.2416228122>.

BATTALY, Heather. Varieties of Epistemic Vice. In: J. Matheson e R. Vitz (Orgs.). *The Ethics of Belief*. Oxford: Oxford University Press, 2014, p. 51-76.

BATTALY, Heather. Epistemic virtue and vice: Reliabilism, responsibilism, and personalism. In: C. Mi, M. Slote, e E. Sosa (Orgs.). *Moral and intellectual virtues in Western and Chinese philosophy*. New York: Routledge, 2016, p. 99-120.

BATTALY, Heather. Testimonial Injustice, Epistemic Vice, and Vice Epistemology. In: I. Kidd, G. Polhaus, e J. Medina (Orgs.). *The Routledge Handbook of Epistemic Injustice*. New York: Routledge, 2017, p. 223-231.

CAIN, William. Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education, in *TechTrends*, 68, 2024, p. 47-57, doi: 10.1007/s11528-023-00896-0.

CASSAM, Quassim. Vice epistemology, in *The Monist*, v. 99, 2016, p. 159-180.

CASSAM, Quassim. Epistemic insouciance, in *Journal of Philosophical Research*, v. 43, 2018, p. 1-20

CAMPOS, Veronica. *Akrasia Epistêmica e Ineficiência Epistêmica: uma abordagem baseada em virtudes*. Belo Horizonte: PPGFil UFMG, 2024.

CLARK, Andy; CHALMERS, David. The Extended Mind, in *Analysis*, v.58, n. 1, 1998, p. 7-19, doi: 10.1093/analys/58.1.7.

CROMPTON, Helen; BURKE, Diane. The Educational Affordances and Challenges of ChatGPT: State of the Field, in *TechTrends* n. 68, 2024, p. 380–392. doi: 10.1007/s11528-024-00939-0.

DAWSON, Shane; JOKSIMOVIC, Srecko; MILLS, Caitlin; GASEVIC, Dragan; SIEMENS, George. Advancing theory in the age of artificial intelligence, in *British Journal of Educational Technology*, v. 54, 2023, p. 1051-1056, doi: 10.1111/bjet.13343.

EAGER, Bronwyn; BRUNTON, Ryan. Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice, in *Journal of University Teaching & Learning Practice*, 20(5), 2023, doi: 10.53761/1.20.5.02.

FUERTES-ALPISTE, Marc. Enmarcando las aplicaciones de IA generativa como herramientas para la cognición en educación Pixel-Bit, in *Revista de Medios y Educación*, n. 71, 2024, p. 42-57, doi: 10.12795/pixelbit.107697.

HILLESUND, Terje. Scholarly reading (and writing) and the power of impact factors: a study of distributed cognition and intellectual habits, in *Frontiers in Psychology*, 17(14), 2023, 1165700, doi: 10.3389/fpsyg.2023.1165700.

JARVELA, Sanna; NGUYEN, Andy; HADWIN, Allyson. Human and artificial intelligence collaboration for socially shared regulation in learning, in *British Journal of Educational Technology*, 54, 2023, p. 1057-1076, doi: 10.1111/bjet.13325

KELLOGG, Ronald. Training writing skills: A cognitive development perspective, in *Journal of Writing Research* 1(1), 2008, doi:10.17239/jowr-2008.01.01.1.

KIRK, Hannah; JUN, Yennie; IQBAL, Haider; BENUSSI, Elias; VOLPIN, Filippo; DREYER, Frederic; SHTEDRITSKI, Aleksandar; ASANO, Yuki. Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. 2021 *35th Conference on Neural Information Processing Systems*, doi: 10.48550/arXiv.2102.04130.

KOBAK, Dmitry; MÁRQUEZ, Rita González; HORVÁT,Emoke-Agnes; LAUSE, Jan. Delving into ChatGPT usage in academic writing through excess vocabulary, in *Science Advances*, 2025, Vol. 11, n. 27, doi: 10.48550/arXiv.2406.07016

KRAEMER, Eric. Virtue Epistemologies and Epistemic Vice, in *Athens Journal of Humanities & Arts*, v.2, 2015, p. 19-30.

KOSMYNA, Nataliya; HAUPTMAN, Eugene; YUAN, Ye Tong; SITU, Jessica; LIAO, Xian-Hao; BERESNITZKY, Ashly; BRAUNSTEIN, Iris; MAES, Pattie. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 2025. Disponível em: arXiv preprint arXiv:2506.08872.

MISHRA, Punya; OSTER, Nicole; HENRIKSEN, Danah. Generative AI, Teacher Knowledge and Educational Research: Bridging Short- and Long-Term Perspectives, in *TechTrends* n. 68, 2024, p. 205-210, doi: 10.1007/s11528-024-00938-1.

RICKMAN, Sam. Evaluating gender bias in large language models in long-term care, in *BMC Medical Informatics and Decision Making*, 25(274), 2025, doi: 10.1186/s12911-025-03118-0.

RIVERA-NOVOA, Angel; DUARTE ARIAS, Daniel. Generative Artificial Intelligence and Extended Cognition in Science Learning Contexts, in *Science & Education*, 2025, doi: 10.1007/s11191-025-00660-1.

RAY, Partha. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, in *Internet of Things and Cyber-Physical Systems*, (3):1, 2023, p. 121-154.

SHARPLES, Mike. Towards social generative AI for education: theory, practices and ethics, in *Learning: Research and Practice*, 9(2), 2023, p. 159-167, doi: 10.1080/23735082.2023.2261131.

SIMMS, Rachells. Work With ChatGPT, Not Against: 3 Teaching Strategies That Harness the Power of Artificial Intelligence, in *Nurse Educator*, 49(3), 2024, p. 158-161, doi: 10.1097/NNE.0000000000001634.

SOSA, Ernest. The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge, in *Midwest Studies in Philosophy*, v. 5, 1980, p. 3-25.

VILAS BOAS, Paulo José; VILAS BOAS, José Vitor. The use of ChatGPT in scientific publishing, in *Geriatric, Gerontology and Aging*, 2023, doi: 10.53886/gga.e0230027.

TENESSINI, Alessandra. Epistemic Vice and Motivation, in *Metaphilosophy*, v. 49(3), 2018, p. 350-367.

ZAGZEBSKI, Linda. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press, 1996.