


Dossiê: Ética e Integridade, Inteligência Artificial, Tecnologias de Informação e Comunicação (TICs) e desinformação nas pesquisas em Humanidades

**Pode a inteligência artificial educar?
Desafios éticos em Inteligência Artificial para uma *tecnopaideia***

**Can artificial intelligence educate?
Ethical challenges in Artificial Intelligence for a *tecnopaideia***

**¿Puede la inteligencia artificial educar?
Retos éticos de la Inteligencia Artificial para una *tecnopaideia***

Mariah Brochado*

 <https://orcid.org/0000-0001-5863-7360>

Resumo: O artigo aborda aspectos técnicos e éticos da área da Computação Cognitiva que simula habilidades cognitivas humanas, notabilizada pelo codinome “Inteligência Artificial” (IA), apontando inconsistências semânticas geradas pela antropomorfização dessas arquiteturas máquinicas, o que reforça a crença na superioridade intelectual de programas computacionais meramente estocásticos. Com apelo propositivo, sustenta-se que é possível conformar essa *praxis* a favor de uma educação tecnologicada que lance mão do aprendizado de máquina (*Machine Learning* – ML) para se obter alto desempenho na formação dos nativos digitais, com práticas que atualizem a clássica noção de *paideia* grega. Adota-se metodologia descritiva e dedutiva, escrutinando criticamente fontes científicas abalizadas sobre o tema, com aportes da Filosofia da Tecnologia. A pesquisa apresenta resultados teóricos relevantes quanto à caracterologia da IA, elidindo equívocos e mitificações sobre suas reais potencialidades, além de trazer para o debate pedagógico a possibilidade de sua adoção para efetivar uma *tecnopaideia* adequada às demandas da era digital.

Palavras-chave: Inteligência artificial. Ética. *Tecnopaideia*.

Abstract: This article addresses technical and ethical aspects of Cognitive Computing, which simulates human cognitive abilities, known by the codename “Artificial Intelligence” (AI), pointing out semantic inconsistencies generated by the anthropomorphization of machine architectures, which reinforces the belief in the intellectual superiority of these purely *stochastic* computer programs. With a propositional approach, it’s argued that it is possible to shape this *praxis* in favor of a technologised education that makes use of Machine Learning (ML) to achieve high performance in the education of digital natives, with practices that update the classical notion of Greek *paideia*. A descriptive and deductive methodology is adopted, critically analysing authoritative scientific sources on the subject, with contributions from the Philosophy of Technology. The research presents relevant results regarding the characterology of AI, eliminating misconceptions and mythifications about its real potential, as well as bringing to the *pedagogical*

* Professora Titular de Filosofia do Direito e da Tecnologia da Faculdade de Direito da Universidade Federal de Minas Gerais (UFMG). Professora Visitante no *Leibniz Institut für Medienforschung – Universität Hamburg*, Alemanha. E-mail: <mbrochado@gmail.com>.

debate the possibility of its adoption to implement a *tecnopaideia* appropriate to the demands of the digital age.

Keywords: Artificial intelligence. Ethics. *Technopaideia*.

Resumen: El artículo aborda aspectos técnicos y éticos del área de la Computación Cognitiva, que simula habilidades cognitivas humanas, conocida bajo la denominación de Inteligencia Artificial (IA), señalando inconsistencias semánticas generadas por la antropomorfización de estas arquitecturas maquínicas, lo que refuerza la creencia en la superioridad intelectual de programas computacionales meramente estocásticos. Con un enfoque propositivo, se sostiene que es posible conformar esta *praxis* a favor de una educación tecnologizada que recurra al aprendizaje automático (*Machine Learning – ML*) para obtener un alto rendimiento en la formación de los nativos digitales, con prácticas que actualicen la noción clásica de la *paideia* griega. Se adopta una metodología descriptiva y deductiva, analizando críticamente fuentes científicas acreditadas sobre el tema, con aportes de la Filosofía de la Tecnología. La investigación presenta resultados teóricos relevantes en cuanto a la caracterología de la IA, al disipar equívocos y mitificaciones sobre sus potencialidades reales; además, incorpora al debate pedagógico la posibilidad de su adopción para hacer efectiva una *tecnopaideia* adecuada a las demandas de la era digital.

Palabras clave: Inteligencia artificial. Ética. *Tecnopaideia*.

Introdução

O desafio primeiro ao se discutir o fenômeno técnico chamado “inteligência artificial” (IA) não é apenas entender o que ela de fato promove, mas colocar sob suspeita o que as pessoas acreditam que ela é capaz de promover. Quando Alan Turing (1950) indagou “Podem as máquinas pensar?”, ele lançou uma provocação que tomou proporções agigantadas nas últimas décadas: a crença generalizada de que máquinas, de fato, podem pensar como humanos. E o uso impreciso, por *experts* da área computacional, de conceitos que já haviam se estabilizado no âmbito das ciências humanas, tais como inteligência, liberdade, autonomia, consciência, passou a suscitar confusões de toda ordem. Ao nos referirmos a esse catálogo terminológico inovador no âmbito da Ciência da Computação, apontamos para uma mudança de sentidos preocupante no que se entende por racionalidade humana, a qual, evidentemente, nos remete às noções de criatividade e liberdade, características que atestam a exclusividade do metabolismo espiritual do *homo sapiens*, como veremos.

O problema de pesquisa suscitado neste artigo consiste, inicialmente, em detectar as falhas conceituais sobre a IA, como medida pedagógica necessária de esclarecimento sobre o que, de fato, significa uma inteligência dita “artificial”. Sobre tal problematização, ergue-se uma segunda, de caráter propositivo: em vez de encetar discursos tecnofóbicos, catastrofizando a adoção das tecnologias consideradas “inteligentes”, em uma tentativa vã de afastar o estado da arte do progresso tecnológico em nossa civilização, seria mais acertado buscar argumentos favoráveis ao uso da IA para desenvolver uma *praxis* educativa condizente com as demandas cognitivas e formativas do século XXI. Para responder a essa questão primordial, três objetivos são apresentados, os quais se desdobram nos quatro tópicos conteudísticos que integram o texto.

O primeiro objetivo é trazer uma caracterologia robusta do fenômeno tecnológico notabilizado por “Inteligência Artificial”, o que será apresentado no primeiro tópico do artigo, intitulado “Caracterização da IA”. A articulação de conceitos dessa área da computação, destinada a indicar independência mecânica no desempenho de máquinas computacionais, tem trazido uma desestabilização inegável para o imaginário social, tanto no que se refere a convicções pacificadas há séculos nas construções culturais dos nômades digitais quanto no que diz respeito à modulação acelerada das práticas da nova geração de nativos digitais.

Tais práticas são marcadas por excessiva insegurança quanto aos limites do humano e por uma dependência crescente quanto a funcionalidades cognitivas, as quais são ofertadas por artefatos cada dia mais miniaturizados e que têm mesmo se tornado extensões da nossa própria corporeidade. Aqui delineamos o segundo objetivo deste ensaio: para além do esclarecimento sobre os significados atribuídos ao desempenho de máquinas digitais, pretendemos desmistificar conceitualmente a apropriação equivocada, pelo senso comum, do que os cientistas da computação consideraram um tipo de desempenho maquínico sofisticado ao ponto de denominá-lo “inteligente”, o que será abordado no tópico seguinte, intitulado “Mitificação sobre a IA”.

Considerando, ainda, que é necessário o exercício crítico sobre as consequências que a usabilidade dessa maquinaria dita “inteligente” vem trazendo para a nossa situação no mundo e para a nossa autocompreensão na nova era digital, importa agora apontar os dois objetivos correlatos aos esclarecimentos trazidos nos tópicos anteriores sobre a ontologia da IA: nos tópicos seguintes, intitulados, respectivamente, “Retidão com IA” e “Educação por IA”, propõe-se uma reflexão sobre como a IA pode ser eticamente adotada para promover educação de qualidade no cenário pedagógico atual, o qual precisa se adequar a novos formatos de aparelhamento didático-tecnológico.

Com intenção propositória, serão trazidos caminhos possíveis para a abordagem adequada a uma educação tecnológica nesse processo vertiginoso de inteiração humano-máquina que se agudiza em nossa era. Em vez de indagarmos, com Turing (1950), “se máquinas podem pensar”, lançamos uma indagação que, de certa forma, conduz esta primeira a nosso favor, qual seja: se máquinas podem pensar, pode a IA, estado da arte da maquinaria “inteligente”, nos ajudar a formular ferramentas didáticas mais adequadas ao ensino dos nativos digitais? O propósito geral deste trabalho é apontar caminhos para essa empreitada na forma de reflexão crítica sobre a IA, carreando conteúdos filosóficos para tanto.

A metodologia adotada é descritiva, quanto ao aporte rigoroso e bem fundamentado de categorizações e conceitos da Ciência da Computação nesta pesquisa, e dedutiva, ao instanciar no texto ora apresentado uma série de decorrências práticas extraídas das reflexões filosóficas que se prestam a ser a senda segura pela qual transitam as críticas e as denúncias originais da pesquisa. As reflexões, predominantemente teóricas, têm um propósito bastante concreto: apontar as inovações possíveis nas práticas pedagógicas herdadas da era analógica, as quais, hoje, contam com um arsenal computacional que garante maior eficiência informativa e formativa com arquiteturas de uma tecnopaideia adequada ao perfil dos nativos digitais.

No estágio de progresso tecnológico a que ascendemos, a pergunta ideal passa a ser: “Pode a IA educar?”. Nossa aposta é de que o auxílio de ferramentas tecnológicas diversas, como o *Scratch* e o *Logo*, por exemplo, e o desenvolvimento de arquiteturas robóticas como agentes morais artificiais (AMAs) em programas pedagógicos, considerando práticas responsáveis inspiradas na Ética Computacional, possa ser um resgate da noção de *paideia* grega, aqui revisitada e atualizada por mecanismos tecnológicos de altíssima *performance* que permitam o melhor desenvolvimento de metodologias compatíveis com as demandas da sociedade digital.

Caracterização da IA

A cognição humana se caracteriza essencialmente por três faculdades: a memorização, o entendimento e o raciocínio. No entanto, os humanos não são dotados de memória imediatamente disponível em tempo real, contendo tudo o que foi apreendido em suas existências. Essa limitação envolve uma outra: não somos capazes de promover análises combinatórias sobre todos os itens

memorizados e, a partir deles, tirar diversas conclusões com precisão e detalhamento, com grande velocidade e eficiência, o que exigiria revisão imediata dos resultados a cada raciocínio formulado.

Para nos substituir em atividades não criativas é que foram arquitetadas máquinas ditas “inteligentes” pela Ciência da Computação, a qual tem por propósito explorar funcionalidades computacionais para maximizar nossas habilidades cognitivas – a Computação Cognitiva¹. E seu feito mais extraordinário recebeu o codinome “Inteligência Artificial” em um ambicioso projeto de John McCarthy, cientista da computação estadunidense, que escreveu uma carta em que aparece a expressão “Inteligência Artificial” pela primeira vez na história da computação². A partir daí, define-se “Inteligência Artificial – IA” como um campo de estudo da Ciência da Computação, a Computação Cognitiva, a qual tem por subcampo o chamado “Aprendizado de Máquina” ou “Aprendizagem Mecânica” (*machine learning*) (Ziviani, 2017). As técnicas dessa área computacional foram inicialmente caracterizadas como “Simulação Computacional”, eis que se destinam a imitar nosso processo cognitivo, com um propósito muito bem definido: produzir resultados a partir da conexão entre ideias (expressas em dados), simulando, assim, a produção de “conhecimento novo”.

O chamado “Aprendizado de Máquina” (AM), por sua vez, são processos computacionais que otimizam a aquisição de informação, exponenciando a execução das tarefas programadas com mais eficiência. Trata-se de um programa de algoritmos que criam outros algoritmos para, segundo o sistema de *feedback*, preencher lacunas informacionais, visando a atingir, a partir dos próprios algoritmos existentes, o desempenho ótimo da máquina. São, portanto, “[...] sistemas de aprendizagem algorítmica capazes de se adaptar a novas situações problemáticas de forma independente e de continuar a escrever seus próprios programas” (Hoffmann-Riem, 2021, p. 15).

Tal ocorre porque os programas de IA vão refinando habilidades a cada vez que repetem as tarefas programadas, em uma sucessiva e crescente aquisição de melhores padrões computacionais, sem necessitar da intervenção imediata dos programadores humanos. É nesse sentido que essa aquisição sem intervenção imediata é chamada de “aprendizado”. Entretanto, nada mais são que programas que atuam sobre programas, permitindo “[...] mudanças adaptáveis no sistema, no sentido de que permitem que o sistema, da próxima vez, faça a mesma tarefa ou tarefas tiradas do mesmo grupo com mais eficiência e eficácia” (Rich; Knight, 1993, p. 514).

Em suma, se uma máquina está municiada de memória e de padrões que operacionalizam os dados dessa memória, similarmente ao que ocorre com a memória e o raciocínio humanos, diz-se que temos aí uma forma de inteligência: a artificial. E, se as tarefas passam a ser executadas sem intervenção imediata de um programador humano, considera-se que há aprendizagem maquina.

Transferir nossas tarefas para máquinas, tal como transferimos a locomoção das nossas pernas para pedais de automóveis e o esforço dos nossos braços para uma lavadora de roupas, liberta-nos das necessidades impostas ao nosso corpo, promovendo crescente bem-estar e longevidade à nossa espécie. Essa “delegação” de tarefas para máquinas é projeção das nossas

¹ “Nesse contexto, é importante refletir sobre os fundamentos a partir dos quais se formou a área da Computação Cognitiva, cuja realização mais exuberante é precisamente a arquitetura da área hoje conhecida como Inteligência Artificial, que, no passado, era simplesmente chamada de Simulação Computacional, do mesmo modo que aquilo que hoje chamamos de computador já foi denominado Cérebro Eletrônico. A ideia de que o cérebro pode ser reproduzido artificialmente e até mesmo emulado em novas formas de inteligência maquina constitui uma nova inspiração para a releitura da tradicional relação homem-máquina, há muito registrada na história da filosofia, particularmente na figura do autômato, desde Aristóteles” (Brochado, 2024b, p. 35, tradução nossa).

² “Nós propomos que um grupo de dez homens realize um estudo de dois meses sobre *inteligência artificial* durante o verão de 1956, no *Dartmouth College*, em Hanover, New Hampshire. O estudo se baseia na ideia de que todo aspecto de aprendizado ou qualquer característica da inteligência consegue, por princípio, ser tão precisamente descrito que uma máquina pode ser criada para simulá-lo” (McCarthy *et al.*, 1955, p. 1, tradução nossa).

próprias habilidades (Kapp, 2015) e atinge desempenho infinitamente superior ao que os nossos membros atingiriam: basta observar o poder de um liquidificador, se comparado ao desempenho da mastigação bucal como estrutura funcional humana.

Os computadores eletrônicos são, na verdade, projeções das habilidades mentais dos humanos, máquinas que conseguem atingir, com maior velocidade e precisão, os mesmos resultados que as atividades conectivas do sistema neural do cérebro humano atingem (Ziviani, 2017). No entanto, como adverte Blay Whitby (2004, p. 38), um filósofo especialista em Ética para a IA da Universidade de Sussex, o computador nada mais é que “um pateta completamente obediente” que executa um algoritmo (padrão de passos simples) para a obtenção de um resultado. Os “[...] cálculos matemáticos podem ser quebrados em passos simples que podem ser executados pelo nosso pateta completamente obediente” (Whitby, 2004 *apud* Brochado, 2023b, p. 294). Por outro lado, os processos eletrônicos funcionam em uma velocidade gigantesca, demorando apenas frações de segundos para se completarem, isto é: “[...] a estupidez tediosa da maneira como os computadores digitais operam é compensada por sua tremenda velocidade” (Whitby, 2004, p. 59-60).

Ocorre que essas conquistas meramente tecno-operacionais são anunciadas ao público leigo de forma não rigorosa tecnicamente, gerando, em torno dos feitos da computação, uma mística que em nada se assemelha aos processos longos e enfadonhos de construção de programas computacionais, levados a cabo por milhares de programadores que há décadas se dedicam a codificar a linguagem humana e imitar nossos processos mentais.

Mitificação sobre IA

Discutir os desafios éticos quanto ao uso da IA é começar por admitir que as habilidades criativas dos nativos digitais estão sujeitas a um processo acelerado e radical de inteiração humano-máquina, o que pode desencadear uma hibridização sem precedentes na história evolutiva da nossa espécie, caminhando, talvez, para um transhumanismo que se revelará contrário ao que identificamos em nossa espécie como transcendência humana.

Atualmente, já se percebe uma forte tendência à negação dessa condição pelas pesquisas neurocientíficas que atestam a determinabilidade dos comportamentos humanos por causações de base neuronal, negando que haja liberdade e consciência tal como acreditamos. Como defende o neurobiólogo da Universidade de Stanford, Robert Sapolsky (2025, p. 32), o monitoramento de milhões de neurônios revela que “[...] quando julgamos estar consciente e livremente decidindo fazer uma coisa, o curso da ação neurobiológica já foi iniciado. Aquela sensação de intenção consciente é uma interpretação tardia e irrelevante” [sobre o que já foi definido por cadeias neuronais]. Essa é uma discussão sobre a existência ou não do livre-arbítrio, travada entre filósofos e neurocientistas, mas que ainda não encontrou uma explicação incontestante sobre a capacidade humana de transcender as circunstâncias vividas e se autorreferenciar ao experimentá-las como substância diversa delas.³

³ “Muito se tem investido na aplicação da potência computacional para simular o cérebro humano, em vez de dissecá-lo em sua totalidade, como ocorre no *Human Brain Project* (HBP), financiado pela União Europeia e parte do projeto suíço *Blue Brain*. Liderado pelo neurocientista Henry Markram, o HBP tem como objetivo simular o cérebro humano em supercomputadores que realizam cerca de seis quatrilhões de operações por segundo. O ponto é que a velocidade computacional e o acúmulo de bilhões de dados não garantem que uma propriedade emergente seja simulável. O modelo computacional mais avançado não é capaz desse feito, porque essas propriedades não decorrem de uma somatória funcional, sem mencionar que, no que diz respeito ao acúmulo de dados, sequer há consenso entre neurocientistas sobre onde, como e em que quantidade as memórias são armazenadas no cérebro” (Brochado, 2024b,

Se por determinismo causal ou não, se por uma força anterior às causalidades biológicas ou não, fato é que existe em nossas experiências uma capacidade de transcender as situações determinadas, situando-nos além delas, dando-lhes sentido e nos referenciando nesse processo significacional sobre a realidade bruta, o que constitui nosso metabolismo espiritual. Esse segundo metabolismo, exclusivo dos humanos, é o que nos caracteriza como seres inteligentes dotados de juízo sobre o bem e o mal (Vaz, 2000, p. 33). E esse metabolismo parece estar sendo atribuído ao desempenho de máquinas consideradas inteligentes.

Não intentamos aqui vilanizar o progresso tecnológico, mas, sim, estimular a reflexão sobre as consequências angariadas pela atribuição de qualificativos humanos a seres máqunicos. Ora, o conforto em se permitir depender (de) e confiar (em) máquinas supostamente superiores a nós em nossas próprias habilidades traz consigo o risco de desespiritualização da nossa própria identidade tal como a concebemos, definida pela intencionalidade e pela responsabilidade sobre as decisões tomadas (Vaz, 2000 *apud* Brochado, 2023a, p. 95).

A dependência de uma IA que falseia a nossa compreensão de mundo e assume protagonismo nas nossas tomadas de decisão, como quando se usa uma ferramenta de IA generativa para escrever textos em substituição à nossa própria capacidade criativa, é uma moldura tecnológica sedutora que nos deixa letárgicos nesse processo de dependência e anulação das nossas próprias competências intelectuais. Esse processo surte um efeito moral devastador: a adesão mecânica a convicções que não são por nós assimiladas e sobre as quais não temos referência, sequer sabemos de onde vêm e que sabidamente são apropriadas anonimamente em milhares de fontes acessadas em rede, é um padrão comportamental oposto ao que nossas crenças indicavam quanto ao respeito à criatividade alheia, às diversas identidades, individualidades criativas que não podem ser dissolvidas em um limbo de ideias referidas a absolutamente nenhum autor.

Sabidamente, estamos nos apropriando de um amontoado de dados que tecnicamente só fazem usurpar e substituir criatividades e negar suas respectivas autorias, lançando em uma placa de transistores tudo o que a nossa civilização construiu. Essa é uma radical alienação da nossa própria identidade intelectual, o que, por esse exercício constante de apropriação de produtos sem identidade, tem nos tornado indiferentes à própria concepção de identidade, tornando-nos apáticos a quaisquer consequências relativas a ações de um sem-número de identidades cerebrais delegadas a um sem-número de estruturas de silício.

Complementarmente, hoje vivemos uma verdadeira mitificação antropomorfizante da IA, o que nos impõe refletir sobre o principal desafio ético da sua utilização: estamos sendo sugestionados a atribuir características humanas ao que não passa de programas computacionais com altíssima capacidade de acumulação e cruzamento de dados por meio de receitas executáveis (algoritmos) para atingir objetivos. O próprio conceito de algoritmo, que nada mais é que um cálculo preciso com variantes delimitadas, vem sendo tomado por essa aura antropomorfizante, de modo que passamos a nos referir a esses processos de cálculos, “os algoritmos”, como verdadeiras entidades à espreita para controlar nossas vidas no imenso e complexo *Big data*. Contudo, de onde brotam essas mitificações sobre entidades técnicas autônomas?

Brien Cantwell Smith, professor de Ciência da Computação e Filosofia da Tecnologia na Universidade de Toronto (falecido em setembro de 2025), registrou que muitos equívocos do entendimento popular sobre os feitos da computação são também fruto de crenças fantasiosas que os próprios profissionais da área compartilham sobre o que esses processos computacionais são capazes de realizar (Smith, 2019). Essa mitificação chega ao ponto de levar alguns cientistas a dedicarem páginas de artigos científicos e jornalísticos à discussão sobre a possibilidade de se

p. 44, tradução nossa).

construir uma “Inteligência Artificial Geral” totalmente independente dos programadores, a qual poderá evoluir para uma “Inteligência Artificial Sobre-Humana” que conseguiria emular nosso cérebro e nossa mente e adquirir consciência, superando-nos e controlando-nos (Boden, 2020, p. 38, 199). No entanto, essas hipóteses são quimeras, já que desenvolver tais programas ainda está longe do que o atual estágio das pesquisas computacionais consegue vislumbrar em curto e médio prazos.

Não é novidade a relação que os humanos estabelecem com objetos inanimados, projetando neles suas próprias características e tratando-os como se algo de humano tivessem. Basta observar uma criança com sua boneca, um violinista com seu instrumento musical ou um colecionador com seu carro antigo dos sonhos. A *agalmatofilia*, atração por estátuas, bonecas etc., já era uma prática comum entre os gregos (Brochado, 2024a, p. 56-57). A extensão dessa prática a máquinas de alta *performance* parece irresistível. Segundo relatam Rich e Knight (1993, p. 665-666), a máquina de lavar louça foi chamada, inicialmente, de “robô” por seus projetistas, pois o que ela realizava era algo diferenciado à época. Entretanto, à medida que seu desempenho deixou de ser surpreendente, o termo “robô” deixou de ser usado e ela foi lançada na vala comum de “simples máquina” executora de uma “tarefa simples”, restando frustrada a mística da “autonomia” de um robô, por hipótese similar a humanos.

Com o advento da Computação Digital e da Robótica, a projeção de habilidades humanas em máquinas ficou ainda mais refinada, a exemplo da famosa humanoide Sophia, que se tornou cidadã saudita.⁴ Assim, a confusão entre máquina do tipo autômato e entidade autônoma passa a ser frequente, o que é potencializado por novos “interlocutores”, *chatbots* que, inclusive, ostentam nomes que sugerem alguma humanidade virtualizada, tais como: Siri (da Apple); Alexa (da Amazon); Bixby (da Samsung); Lu (da Magazine Luiza); Bia (do Bradesco); e também Victor, Maria e Rafa (sistemas de IA do Supremo Tribunal Federal).

Neste ponto, é imperioso ressaltar que os cientistas da computação são bastante pragmáticos ao adotarem analogias para indicar os feitos maquímicos, como ocorre com o emprego das palavras “inteligência”, “aprendizagem” e “autonomia”, pouco se importando com a precisão semântica a elas atribuída pela Filosofia. Um bom exemplo pode ser encontrado na obra seminal de Alan Turing (1950), *Computing machinery and intelligence*, na qual ele lança a provocação “podem máquinas pensar?” e consta um tópico intitulado “Máquinas que aprendem”. Nesse texto, Turing defende que máquinas “inteligentes” devem ser projetadas como crianças que podem aprender com as informações adquiridas. Para tanto, é necessário submeter essas máquinas a um aprendizado como “[...] um percurso educativo similar ao humano, pois supostamente nosso cérebro é parecido com um bloco de notas cheio de papéis em branco e com pouquíssimos mecanismos” (Turing, 1950 *apud* Brochado, 2023b, p. 227). Trata-se de uma visão mecanicista do organismo humano, o qual poderia ser projetado em máquinas.⁵

Outro exemplo é o conceito que John McCarthy propõe para “livre-arbítrio”, entendendo-o como uma mera capacidade funcional relacional que pode ser decalcada em máquina. Ele defende que humanos decidem o que fazer ao avaliarem os resultados obtidos por um processamento de alternativas, segundo as diversas ações por eles executadas. Esse proceder que avalia resultados a partir de experiências bem e malsucedidas pode ser reproduzido em máquina, com uma vantagem: nesta, o processo ocorrerá de forma determinística, o que é um ganho em experiência, por acúmulo,

⁴ Disponível em: <https://www.instagram.com/realsophiarobot?igsh=NnQzZHFqbTV2bDR2>. Acesso em: 9 maio 2026.

⁵ A propósito das teses mecanicista e vitalista, ver Brochado (2023b, p. 128-145).

e evolução, por aprendizado em *feedback*, sem os desvios típicos da liberdade humana (McCarthy; Hayes, 1969, p. 6, 2).

É essencial compreendermos que a área da computação trata nossas habilidades como procedimentos que alcançam resultados, o que nada tem a ver com as concepções metafísicas sobre a natureza humana. Diante disso, somos levados a indagar quais limites éticos devem ser trazidos para esse campo científico, questionando apropriações conceituais levianas que podem macular o bem-estar que a inovação tecnológica deve trazer para a sociedade, comprometendo nossas conquistas civilizatórias.

Em termos epistemológicos, estamos propondo aqui um exercício crítico sobre um fenômeno técnico que vem sendo confundido com experiências subjetivas, configurando um *mix* aparentemente espúrio entre sujeito e objeto. Esta discussão exige a presença de outro nível de refinamento das reflexões, o que acreditamos ser próprio dos estatutos epistemológicos da Filosofia da Tecnologia, e não da própria Ciência da Computação. Esta, como ciência que é, ocupa-se dos fundamentos técnicos da IA, e não da crítica humanista sobre o sentido (*ratio*) e os fins (*telos*) da IA, ou seja, seus fundamentos éticos.

Dito isso, nossa tarefa agora deve ser menos conjectural quanto a um futuro escatológico submetido a uma hipertecnologia e mais propositiva quanto ao que o presente nos sinaliza sobre as possibilidades de uso ético da IA para exponenciar nossas próprias habilidades e nosso bem-estar.

Retidão com IA

Alguns autores sugerem que uma área de aplicação da Filosofia da Tecnologia é a “Ética Computacional”. Conforme caracteriza Moor (1985), a Ética Computacional se dedica a analisar a natureza da tecnologia computacional e seus impactos sociais, o que envolve a formulação fundamentada de políticas para o uso ético dela decorrentes.

O autor justifica a adoção da expressão “tecnologia computacional” por compreender que suas reflexões devem abranger todas as tecnologias associadas ao desempenho da máquina computador, tais como as questões relativas a *software* e *hardware*, à rede mundial de computadores, às novas habilidades propiciadas pela inteiração dos humanos com computadores, às ações delas emergentes (tanto no plano individual quanto no plano social) e ao vazio político existente sobre “como tais tecnologias devem ser utilizadas”. É tarefa fundamental da Ética Computacional determinar políticas para orientar nossas ações para o uso ético da tecnologia da computação. Moor (1985) defende que se trata de uma área de pesquisa complexa e dinâmica que deve relacionar a computação com novos valores que dessa tecnologia emergem, devendo ser considerada não apenas um campo entre ética e ciência, mas “[...] uma disciplina por direito próprio, que fornece tanto conceituações para a compreensão quanto políticas para o uso da tecnologia da computação” (Moor, 1985, p. 268).

No âmbito da ética aplicada à formulação de políticas destinadas a boas práticas tecnológicas, é possível refletir sobre a possibilidade de conversão da IA a nosso favor, indo na contramão do risco acima sugerido sobre uma possível desespiritualização do humano. A reflexão que propomos é: podemos usar a IA para exponenciar nossa evolução moral em vez de crer que ela irá nos mediocrizar em nossas habilidades até nos alienar da nossa própria essência espiritual? Nesse cenário, podemos prospectar novos *designs* de programas de IA que nos estimulem, com sua peculiar precisão computacional, a evoluir e alcançar a tão almejada retidão moral que nos leve a agir por identificação com sugestões criativas rumo à prática do bem, do justo, do caritativo *per*

quandam connaturalitatem (Vaz, 1999, p. 469) moral? Podem os programas de IA se tornar *experts* em estímulo empático às ações virtuosas?

Ora, se a máquina atinge maior precisão do que os humanos na escolha dos melhores desempenhos na prática de qualquer tarefa, uma questão ética que se coloca é: podemos ser melhorados moralmente a partir de programas pedagógicos desenvolvidos por computadores? Como vimos, o chamado aprendizado de máquina é um programa de algoritmos que criam algoritmos para, segundo o sistema de *feedback*, preencher lacunas informacionais, visando a atingir, a partir dos próprios algoritmos existentes, o desempenho ótimo da máquina. A partir desse sistema de aprendizado, é possível estabelecer base de dados com conteúdo moral para imitar a moralidade humana, tal como já conseguimos imitar as conexões da inteligência humana, às quais chamamos de raciocínio?

Certo é que a lista da Filosofia Clássica sobre virtudes cardeais e teologais já consta de bases de dados que foram produzidos por séculos, tal como encontramos na obra aristotélica e nos roteiros para a vida virtuosa dos estoicos. O desafio que lançamos é técnico e ético: seria possível resgatar os princípios da Ética das virtudes em programas atualizados de IA que possam ser usados para ensinar boas práticas aos humanos? Estaríamos lidando com um modelo comportamental maquínico de falibilidade controlada quanto a recalitrâncias (típicas dos humanos) e em constante evolução (pelo aprendizado de máquina).

O agenciamento maquínico computacional em linhas de desterritorialização, não mais como conteúdo, mas como força e função (Deleuze; Guattari, 2011, p. 233-234), pode ser um instrumento para realizar tarefas de natureza imediatamente ética, visando diretamente ao nosso processo formativo (*paideia*). Parece ser bastante razoável que a mencionada antropomorfização⁶

⁶ Trazemos esta nota de rodapé para registrar a advertência da *Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente*, adotada pela Comissão Europeia para a Eficiência da Justiça (CEPEJ) em sua 31ª reunião plenária (Estrasburgo, 3 e 4 de dezembro de 2018), quanto à antropomorfização da IA e à ingênua crença em IA forte. Não é esse o escopo do artigo, mas é importante trazer esse relevante registro jurídico, ainda que se trate de *soft law*. Assim está disposto na Carta Europeia:

“62. Em termos gerais, também é importante ter em mente a noção antropomórfica de que as máquinas de computação são inteligentes e que seus projetistas conseguiram deslizar uma mente dentro dos seus mecanismos. Infelizmente, essa ideia ainda permeia muitas análises de justiça preditiva que emprestam a esses dispositivos capacidades imediatas ou futuras para a replicação próxima da inteligência humana. Este contexto, alimentado todos os dias por uma nova série de avanços revolucionários da IA, leva-nos a todos a abordar estas ferramentas preditivas com uma certa dose, consciente ou não, de misticismo, afirmando por vezes que aquilo que não é inteiramente possível hoje será inevitavelmente possível amanhã.

63. As IAs ‘fortes’ da literatura de ficção científica não existem. Este tipo de IA, que seria dotada não só de inteligência, mas também de consciência, permanece puramente ficcional. Os sistemas de autoaprendizagem que estão a ser desenvolvidos atualmente são descritos como IAs ‘fracos’ e são capazes de extrair padrões complexos e aprender com grandes volumes de dados de forma eficiente e muitas vezes com altos níveis de precisão preditiva.

64. Para superar quaisquer considerações instintivas ou fantasiosas, estas tecnologias de processamento e análise de informação devem ser compreendidas com base nos três conceitos seguintes:

- IA não é um objeto único e homogêneo: ele é, na verdade, um conjunto de ciências e técnicas (matemática, estatística e ciência da computação) capaz de processar dados para projetar tarefas de processamento computacional muito complexas.

- Os motores da IA não produzem inteligência por si só, mas usam uma abordagem indutiva: a ideia é associar de forma quase automatizada um conjunto de observações (entradas) com um conjunto de resultados possíveis (saídas), usando várias propriedades pré-configuradas. Especificamente para a justiça preditiva, o motor constrói ligações entre os diferentes grupos lexicais que compõem as decisões judiciais. Estes grupos estão correlacionados entre os identificados na fase de entrada (fatos e raciocínio) e os identificados na fase de saída (a parte operacional da decisão) e depois classificados.

- A confiabilidade do modelo (ou função) construído depende fortemente da qualidade dos dados utilizados e da

crescente nas relações com programas de computador (inclusive nomeando-os como se interlocutores fossem) deva ser direcionada para formar base de dados que estimulem a empatia pelas práticas virtuosas nas relações humanas, em vez de nos alienar da convivência humana ao nos encerrar em uma vida pautada por “telas”. Tais práticas devem trazer esclarecimento sobre a real natureza maquínica, evitando-se uma dependência que, atualmente, tem caminhado para duas distorções perigosas:

- i) a primeira é o risco que assumimos ao transferir nosso agir (livre e responsável) para máquinas, tendendo ao processo que os filósofos da Ética chamam de “cloroformização” confortável da nossa consciência (Brochado, 2021, p. 156). O ato de consciência, que implica ser testemunha (*conscius*) dos próprios atos, responsabilizando-se pelas escolhas eleitas, vai se estagnando letargicamente ao transferir para os infundáveis recursos maquínicos nossas tomadas de decisão, substituindo-nos de forma banalizada em nosso cotidiano;
- ii) a segunda é o já mencionado processo crescente de “desespiritualização” da nossa identidade moral (Vaz, 1992 *apud* Brochado, 2021, p. 89), visto que a prática reiterada de transferência decisional para programas computacionais reconhecidos como entidades similares (e até superiores) a humanos nega ao humano sua constituição primordial, seu metabolismo espiritual, tal como definido acima.⁷

Qualquer programa de Ética Computacional deve ser arquitetado no sentido contrário a essas duas vias que desumanizam o humano ao manter o obscurecimento sobre os limites de qualquer máquina, já que nenhuma delas é de fato inteligente.

É urgente um exercício maiêutico que leve os jovens usuários de artefatos computacionais a refletirem sobre a possibilidade de que as ferramentas de IA sejam apenas versões mais lúdicas e convidativas ao aprendizado de boas práticas integradas à convivência escolar. No Brasil, parece que estamos indo no sentido contrário a essa compreensão de literacia digital, ao ter sido promulgada a Lei nº 15.100, de 13 de janeiro de 2025. Nela, está previsto: “Fica proibido o uso, por estudantes, de aparelhos eletrônicos portáteis pessoais durante a aula, o recreio ou intervalos entre as aulas, para todas as etapas da educação básica” (Brasil, 2025, art. 2º). A imposição de não uso, sem se fazer acompanhar por investimentos em projetos pedagógicos inclusivos de ferramentas tecnológicas avançadas, não parece ser a melhor alternativa para a formação (*paideia*) dos nativos digitais. No entanto, essa é outra questão, que não será debatida neste artigo, embora mereça especial atenção em um projeto de *tecnopaideia*, como formação municiada de tecnologias avançadas próprias do nosso tempo, a era digital.

Para unir o desafio de formular pedagogias inclusivas com máquinas inteligentes aos estatutos da Ética filosófica, é possível que uma *digital literacy* formulada no campo da Ética Computacional possa resgatar o aristotelismo, por exemplo, buscando na Ética das virtudes o caminho para criar máquinas inteligentes na forma de “agentes morais artificiais (AMAs)”. Vejamos.

Educação por IA

Agentes morais artificiais (AMAs) podem ser desenvolvidos para funcionar como modelo comportamental para os humanos, como seres “virtuosos” infalíveis que essas máquinas podem se

escolha da técnica de aprendizagem automática” (CEPEJ, 2018).

⁷ Ver Vaz (2000, p. 237).

tornar com a precisão da Algoritmia e a constante atualização/complementação por retroalimentação do Aprendizado de Máquina. Esses AMAs teriam por tarefa ensinar boas práticas aos humanos, seja partindo dos estatutos da Ética das virtudes, seja da Ética dos valores. Ambos são bons modelos para catalogações, considerando que, tecnicamente, as orientações comportamentais devem ser bem definidas para serem programas de *software*, para o que é necessário especificar cada detalhe da ação escorregia a ser praticada com vistas ao **resultado** pretendido.

Em qualquer caso, a base empírica comportamental é essencial para a Ciência da Computação, já que o ponto de partida para a programação desses agentes morais artificiais são comportamentos observáveis e replicáveis, devendo ser muito bem discriminados no esquema algorítmico “se... então”. Por exemplo: “**se** alguém me dirige uma pergunta sobre a verdade de um fato por mim conhecido, **então** devo revelar a verdade sem distorções quanto aos seus detalhes”. Contudo, essa assertiva é demasiadamente abstrata; ela deve ser discriminada detalhadamente, com base em cada item de fato e na valoração consequencial dele – veja como é difícil a tarefa de “programar” comportamento moral!

Nesse sentido, Erez Firt (2024) argumenta que a extração de valores humanos pode ser realizada pela observação de indivíduos cuidadosamente selecionados, considerando as recompensas almeçadas, o que pode ser executado pelo método de aprendizado maquínico por reforço inverso, segundo o qual o sistema de IA tenta descobrir as recompensas que os sujeitos em observação estão tentando otimizar.

Aquí propomos uma nova leitura maquinizada da Ética aplicada, partindo da adoção do rigorismo eficaz da IA para atingir precisão ótima na prática de ações virtuosas. Evidentemente, a máquina não tem capacidade moral para julgar se ações são boas ou más, justas ou injustas, pois o que os cientistas chamam de aprendizado é apenas um sistema de acúmulo de dados sem autojulgamento sobre cada dado apreendido. Não é disso que se trata. Não temos por propósito discutir se máquinas têm vontade, o que foi superado de largada nas primeiras páginas deste ensaio.

Voltando a Cantwell Smith (2019), ele refuta com muita propriedade a conjectura de que inteligência **artificial** seja inteligência **genuína**, pois ela não emite juízo sobre as informações que recebe: não só quanto às fontes de recepção, como também quanto à forma como os dados são integrados, em diversas perspectivas, de modo responsável, face ao mundo a eles subjacente (Smith, 2019). Essa concepção de responsabilidade como autorreferenciamento quanto às escolhas eleitas é o que chamamos “intencionalidade”, o que, evidentemente, placas de silício não são capazes de promover tecnicamente. Para Smith (2019), máquinas só são capazes de *reckoning*, jamais de *judgment*, quer dizer, o que elas realizam é uma espécie de “racionalidade calculadora”, pois manipulam representações que são semanticamente interpretáveis, mas esses sistemas de interpretação não são capazes de compreender o que são essas representações. E não o são por não serem dotados de capacidade de se responsabilizarem pelo conteúdo de suas próprias representações, precisamente por não serem capazes de se envolver autenticamente com a forma de ser do mundo no qual suas representações se dão.

Em contraposição, diz Smith (2019), ter *judgment* é ser capaz de tomar objetos por objetos, os quais são distintos de sujeitos; é ser apto a distinguir aparência e realidade; é estar existencialmente comprometido com sua própria existência e com a integridade do mundo, entendendo-se como diferença destacada neste mundo, entre objetos que limitam existências (Smith, 2019, p. 110). A racionalidade calculadora não tem prudência, bom senso, como se diz vulgarmente, e a razão não é só cálculo, como creem alguns pesquisadores da área de IA. Ser capaz de julgamento é ter plenamente em conta as consequências dos registros, e não apenas registros. Se qualquer sistema de captação de dados ainda não foi capaz de exercer julgamentos sobre o que

“apreende”, evidentemente, a responsabilidade pela captação e pela integração legítima dessa datificação será sempre e exclusivamente humana, o que nos garante que especialistas em *digital literacy* serão constantes atualizadores de agentes inteligentes de alto quilate pedagógico, alimentando esses programas com conteúdos que serão recorrentemente ajustados para seu melhor desempenho (propósito do *machine learning*).

Fazendo aqui um exercício de ideação para um *software* de AMA que ensine ações virtuosas, temos como ponto de partida a certeza de que um robô programado para praticar condutas para o bem, ao contrário dos humanos, jamais recalcitrará com o padrão moral programado. Trata-se de uma programação precisa, na forma de IA, da moralidade universal e incorruptível idealizada nos tratados de Ética. Com controle e revisão algorítmica constante dos valores adequados a cada fase de desenvolvimento dos estudantes, teremos padrões mais aperfeiçoados do que seria possível em qualquer prática espontânea da racionalidade humana de um ser humano real, falível. Vale dizer: a criação de AMAs funcionaria como uma reabilitação artificializada universal-concreta (efetiva) da razão prática, alcançando práticas virtuosas com excelência e precisão incorruptíveis – ao contrário do que acontece com a *praxis* humana empírica espontânea. A psicométrica computacional e a terapia computacional já experimentam possibilidades dessa natureza ao desenvolverem e utilizarem *softwares* para o desenvolvimento intelectual e amadurecimento psíquico dos usuários. Vejamos alguns exemplos.

Fong, Nourbakhsh e Dautenhahn (2003, p. 146) defendem que um robô para interação social deve ser uma máquina persuasiva, pois ele é usado para alterar o comportamento humano, as atitudes e os sentimentos daqueles com quem interage, tal como se verifica na terapia do autismo. Também é possível usar o robô como avatar quando ele funciona como representação humana a distância. Uma preocupação da robótica epigenética, que desenvolve sistemas que imitam o desenvolvimento biológico prolongado, é a evolução nas habilidades robóticas de acordo com a interação com os usuários no decorrer do tempo. Um robô de estimação pode ser aprimorado para se adequar ao desenvolvimento da criança, mantendo o interesse dela em cada fase do seu desenvolvimento.

Um bom exemplo de interação pedagógica humano-máquina é o aplicativo *Scratch*, que tem sido muito eficiente nos processos educativos, pois seu *design* lúdico estimula a imaginação e a criatividade dos usuários por meio da interatividade com personagens e cenários programados para envolver crianças e jovens (Leal; Aita; Silva, 2020, p. 421-422). Outra proposta interessante, criada em 1967 (portanto, já é bem antiga) com inspiração no construtivismo de Piaget, é a linguagem de programação *Logo*, que usa uma tartaruga gráfica na tela do computador para guiar a formulação de desenhos complexos com comandos simplificados.⁸

⁸ Conforme destacam Jaramillo, Andrade e Caraguay (2024, p. 52), citando outros autores, “[...] la programación como el Logo, el proceso de construcción del conocimiento programático puede ser imaginado y representado de manera más natural, lúdica, con pensamientos propios, con esquemas de secuenciación, de propiedad, de condición, de repetición, de formas estructuradas, organizadas, autoorganizadas”. De igual forma: “Logo muestra un entorno gráfico en el que un objeto con forma de ‘tortuga’ se mueve por toda la pantalla, tiene como característica que las instrucciones se pueden comprender en diferentes idiomas, por lo tal razón, esto hace que Logo sea un lenguaje de programación fácil de aprender” (Jaramillo; Andrade; Caraguay, 2024, p. 28). Exemplos práticos do uso da programação na aprendizagem podem ser encontrados nesse mesmo trabalho de Jaramillo, Andrade e Caraguay (2024), intitulado *Herramientas digitales para la enseñanza de la programación en el ámbito educativo*. Outra referência prática, com êxitos relevantes, é o projeto “ProFuturo”, coordenado pelo chileno Rodrigo Fábrega, que se propõe a desenvolver “[...] experiências inovadoras de ensino-aprendizagem digital para promover o desenvolvimento das competências digitais, permitindo que professores, alunos e diretores das escolas enfrentem os desafios do século XXI” (Fundação Telefônica Vivo; Fundación “La Caixa”, 2026).

O aprendizado de máquina hoje já possibilita o *design* de assistentes pessoais artificiais que manifestem empatia ao interagir com humanos na forma de robôs corporativos destinados à mediação em *compliance* empresarial, ou como robôs babás programados não só para cuidar, mas também para ensinar princípios morais a crianças em fase de formação; ou, ainda, para performar como cuidadores artificiais capazes de interagir dialogicamente com idosos solitários, prestando-lhes suporte físico e emocional (Brochado, 2023a, p. 88).

O *design* robótico pode ser muito variado, incorporando morfologias antropomórficas, zoomórficas e até mesmo caricaturizadas – afinal, animadores há muito tempo comprovam que um personagem não precisa parecer realista para ser crível; e caricaturas podem ser usadas para criar vieses de interação desejados e concentrar a atenção em características específicas do robô (Fong; Nourbakhsh; Dautenhahn, 2003, p. 150). A tendência, inclusive, é que a manifestação de emoções seja o forte da robótica, considerando que há provas de que humanos tendem a tratar robôs como semelhantes, o que fica evidenciado em jogos e agentes de *software* (Fong; Nourbakhsh; Dautenhahn, 2003, p. 151). O *Kirobo Mini*, da Toyota, por exemplo, foi apresentado no Japão como um “androide de companhia”. Ele vem atender a um público vasto, especialmente jovens entre 15 e 39 anos que se enclausuram em casa, recusando-se a ter vida social – chamados *Hikikomori*, literalmente, “isolados em casa” (Han, 2018, p. 29).

Conforme pondera Blay Whitby (2010), é provável que babás robóticas se sofisticem muito, principalmente no Japão, em Cingapura e na Coreia, pois esses países têm atualmente uma vantagem tecnológica significativa na construção de robôs cuidadores, os quais são muito bem aceitos culturalmente. Entretanto, ele adverte que o excessivo uso de babás robóticas pode ser prejudicial psicologicamente à criança, pois, ainda que os pais delimitem o tempo de auxílio do assistente maquínico, em uma cultura de extensa jornada de trabalho, crianças podem ficar mais fortemente ligadas à cuidadora robô que aos pais, e “[...] a capacidade dos pais de ver e ouvir a partir do ponto de vista do robô através do seu celular irá, sem dúvida, ajudar a criar a impressão de que a situação é perfeitamente segura” (Whitby, 2010, p. 293, tradução nossa). Veja que há muitas questões a serem trazidas a debate no campo da pesquisa em Ética aplicada à Robótica, que ainda está engatinhando no Brasil.

Em todo caso, com limitações técnicas e éticas que se impõem, um desafio promissor no atual estágio da civilização da técnica é pensar a relação humano-máquina como forma de aprendizado moral por meio da técnica, ao lançar mão de novas práticas pedagógicas inclusivas da avatarização e da robotização de entidades artificiais moralmente bem arquitetadas com ações convidativas à prática do bem. Tais arquiteturas vêm preencher uma lacuna do mercado tecnológico de ponta, considerando que políticas devem ser pensadas para esse fim, tal como sugere a Ética Computacional trazida por Moor (1985). Sob o ponto de vista filosófico, eis que uma improvável e imprevisível relação aparentemente paradoxal (a técnica ensinar ética) começa a se estabelecer com o uso da precisão computacional. E quiçá máquinas performáticas com seus circuitos em silício passem a ser instrumento para corrigir os desacertos de caráter para a evolução da alma humana.

O aprendizado maquínico virtuoso talvez conduza os humanos a patamares morais mais elevados do que supúnhamos, considerando-se que algoritmos devem representar ações simples e detalhadas, o que exigirá mais cuidado e rigor com a programação de didáticas modelares sobre ações virtuosas. Algoritmos que definam ações virtuosas serão mais precisos e convincentes a cada *feedback*, possibilitando que os próprios humanos assimilem melhor o “passo a passo” que produz

⁹ Disponível em: <https://youtu.be/Ol5VoC9Feq?si=euDmsJrilBPcIcIF>. Acesso em: 9 maio 2026.

o resultado virtuoso. Assim, as razões para a prática do bem estarão mais especificadas, logo, mais compreensíveis.

Trata-se de uma execução atualizada da Ética Intelectualista clássica, cujo princípio maior é: só pratica o mal quem *não* detém pleno conhecimento do bem (Vaz, 1999), haja vista que seria uma *apaidensia* concluir que uma ação é boa e negá-la pela prática contrária a ela. O que defendemos aqui é, portanto, a possibilidade de resgate da formação do homem grego, uma *paideia*, por assim dizer, analógica, agora conformada aos padrões refinados de uma *tecnopaideia* para nativos digitais.

As bases da Ética Finalística seriam, enfim, depositadas em programas com precisão conceitual, a partir de observações de dezenas de comportamentos e de seus resultados. E o aprendizado das virtudes seria otimizado por programas customizados em avatares e gamificações atrativos para crianças e jovens que vivem a interface constante com as telas. Com o uso da IA na forma de agentes avatarizados e/ou gamificados, com *design* apropriado à *digital literacy*, teríamos usabilidade eticamente saudável da computação para auxiliar pais e professores. Não é idealismo exacerbado apostar que o intelectualismo ético grego possa ser revisitado como expressão máxima de uma (tecnó)*paideia* que reabilita, por meio de máquinas computacionais exemplares em suas bases de dados, a Ética clássica da virtude-ciência (Brochado, 2023a; Vaz, 1988).

O pano de fundo comum a todas as questões aqui trazidas é certamente o nosso maior desafio ético: compreender a relação humano-máquina como mecanismo de aprendizado em prol da nossa evolução intelectual e moral. Ao contrário do que supõem os tecnofóbicos mais radicais, o binômio “maquinização humana/humanização maquínica” talvez seja o prelúdio de um (trans)humano mais evoluído moralmente e amadurecido espiritualmente.

Conclusão

O primeiro passo para uma boa prática quanto à usabilidade ética de ferramentas de IA é desmistificar a natureza dessa simulação computacional dita inteligente. Longe de ser uma entidade consciente ou autônoma, a IA permanece uma sofisticada ferramenta de simulação de nossas próprias habilidades cognitivas, otimizada para tarefas específicas, mas fundamentalmente um “pateta obediente”, como adverte Whitby (2010).

A persistente tendência à antropomorfização, alimentada por uma semântica técnica imprecisa que estimula o fascínio popular sobre máquinas “inteligentes”, representa um dos maiores riscos éticos atualmente, pois obscurece a real natureza maquínica da IA e fomenta uma dependência constante de artefatos digitais. A dependência de uma IA que falseia a nossa compreensão de mundo e assume protagonismo em nossas tomadas de decisão é o caminho distorcido do bom uso dessas ferramentas de IA, prestando-se, no fundo, a degradar a nossa identidade intelectual e moral, tornando-nos indiferentes quanto às consequências das nossas decisões.

Não restam dúvidas de que a reflexão filosófica sobre a tecnologia se torna chave mestra nesta nova versão da inteiração humano-máquina, não para colocar em xeque a importância da inovação tecnológica que promove bem-estar, mas para questionar o sentido (*ratio*) e os fins (*telos*) da IA. Seguindo essa convicção, o artigo aventa a possibilidade de reverter o risco de alienação sobre o uso da IA, propondo o uso ético e pedagógico de ferramentas que se alinhem ao ideal de *paideia* como formação humana integral, o que pode ser articulado com práticas pedagógicas inclusivas de agentes morais artificiais (AMAs).

O uso de assistentes robóticos pode estimular o discernimento sobre os usos benéficos da IA no campo pedagógico, seu potencial transformador e adequado ao desenvolvimento intelectual e ético dos humanos, garantindo, assim, que a tecnologia sirva à evolução do espírito, e não ao seu amedrontador amesquinamento, como vem sendo aventado pelos tecnofóbicos mais radicais.

Referências

Boden, M. A. **Artificial intelligence**: a very short introduction. 2. ed. Oxford: Oxford University Press, 2020.

BRASIL. Lei nº 15.100, de 13 de janeiro de 2025. Dispõe sobre a utilização, por estudantes, de aparelhos eletrônicos portáteis pessoais nos estabelecimentos públicos e privados de ensino da educação básica. **Diário Oficial da União**: seção 1, Brasília, DF, n. 9, p. 3, 14 jan. 2025.

BROCHADO, M. **Ética e direito**: pelas trilhas de Padre Vaz. Curitiba: CRV, 2021.

BROCHADO, M. Inteligência artificial e ética: um diálogo com Lima Vaz. **KRITERION**, Belo Horizonte, n. 154, p. 75-98, abr. 2023a. DOI: <https://doi.org/10.1590/0100-512X2023n15404mb>

BROCHADO, M. **Inteligência artificial no horizonte da filosofia da tecnologia**: técnica, ética e direito na era cibernética. São Paulo: Editora Dialética, 2023b.

BROCHADO, M. Die Anthropomorphisierung der Künstlichen Intelligenz: ein notwendiger Dialog zwischen Philosophie der Technologie und Rechtsphilosophie. In: BROCHADO, M. (org). **Direito e Estado entre mundo analógico e era digital**: reflexões de fronteira em homenagem a Wolfgang Hoffmann-Riem. São Paulo: Editora Dialética, 2024a. p. 55-79.

BROCHADO, M. Epistemological and ethical reflections on artificial intelligence as an Area of cognitive computing. **Great Britan Journals Press**, Londres, v. 24, n. 6, p. 33-52, 2024b.

COMISSÃO EUROPEIA PARA A EFICIÊNCIA DA JUSTIÇA. Carta Europeia de Ética sobre o Uso de Inteligência Artificial em Sistemas Judiciais e Seu Ambiente. **CEPEJ**, dez. 2018. Disponível em: <https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0>. Acesso em: 9 maio 2026.

DELEUZE, G.; GUATTARI, F. **O anti-édipo**: capitalismo e esquizofrenia. São Paulo: Editora 34, 2011.

FIRT, E. Ought we align the values of artificial moral agents? **AI and Ethics**, [s. l.], v. 4, p. 273-282, 2024. DOI: <https://doi.org/10.1007/s43681-023-00264-x>

FONG, T.; NOURBAKHSH, I.; DAUTENHAHN, K. A survey of socially interactive robots. **Robotics and Autonomous Systems**, [s. l.], v. 42, p. 143-166, 2003. DOI: [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)

FUNDAÇÃO TELEFÔNICA VIVO; FUNDACIÓN “LA CAIXA”. Quem somos. **ProFuturo**, [s. l.], 2026. Disponível em: <https://profuturo.education/pt-br/quem-somos/>. Acesso em: 9 maio 2026.

HAN, B.-C. **No enxame**: perspectivas do digital. Tradução: Lucas Machado. Petrópolis: Vozes, 2018.

HOFFMANN-RIEM, W. **Teoria geral do direito digital**. Transformação digital, desafios para o direito. São Paulo: Forense, 2021.

JARAMILLO, M. L.; ANDRADE, M. A. C.; CARAGUAY, G. M. **Herramientas digitales para la enseñanza de la programación en el ámbito educativo**. Ciudad Universitaria Guillermo Falconí, Loja, Ecuador: Editorial Universitaria, 2024. Disponível em: https://unl.edu.ec/sites/default/files/archivo/2024-12/Herramientas%20digitales%20para%20la%20enseñanza%20de%20la%20programación%20e%20el%20ámbito%20educativo_1.pdf. Acesso em: 10 maio 2026.

KAPP, E. **Grundlinien einer Philosophie der Technik: zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten (1877)**. Hamburg: Felix Meiner Verlag, 2015.

LEAL, L. B.; AITA, K. M. de S. U.; SILVA, A. M. L. Terapia Computacional: uma vivência voltada para o tratamento de pacientes com transtornos psíquicos mediado pela ferramenta Scratch. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – CBIE, 9., WORKSHOP DE INFORMÁTICA NA ESCOLA, 26., 2020, Natal. **Anais eletrônicos** [...]. Natal: SBC, 2020. Disponível em: <https://sol.sbc.org.br/index.php/wie/article/view/12634/12501>. Acesso em: 2 mar. 2026.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. A proposal for the Dartmouth summer research project on artificial intelligence. **mmcgull**, [s. l.], 31 ago. 1955. Disponível em: <https://tinyurl.com/24oh6yr4>. Acesso em: 10 maio 2026.

MCCARTHY, J.; HAYES, J. P. **Some philosophical problems from the standpoint of artificial intelligence**. Stanford: Stanford University, 1969.

MOOR, J. H. What is computer ethics? **Metaphilosophy**, v. 16, n. 4, p. 266-275, 4 out. 1985.

RICH, E.; KNIGHT, K. **Inteligência artificial**. Tradução: Maria Cláudia Santos Ribeiro Ratto. São Paulo: Makron *Books*, 1993.

SAPOLSKY, R. M. **Determinados: a ciência da vida sem livre arbítrio**. Tradução: Berilo Vargas. São Paulo: Companhia das Letras, 2025.

SMITH, B. C. **The promise of artificial intelligence: reckoning and judgment**. Cambridge: MIT Press, 2019.

TURING, A. M. Computing machinery and intelligence. **Mind**, New Series, v. 59, n. 236, p. 433-460, out. 1950.

VAZ, H. C. L. **Escritos de filosofia II: ética e cultura**. São Paulo: Edições Loyola, 1988.

VAZ, H. C. L. **Escritos de filosofia IV: introdução à ética filosófica 1**. São Paulo: Edições Loyola, 1999.

VAZ, H. C. L. **Escritos de filosofia V: introdução à ética filosófica 2**. São Paulo: Edições Loyola, 2000.

WHITBY, B. **Inteligência artificial: um guia para iniciantes**. Tradução: Claudio Blanc. São Paulo: Madras, 2004.

WHITBY, B. Oversold, unregulated, and unethical: Why we need to respond to robot nannies. **Interaction Studies**, [s. l.], v. 11, n. 2, p. 290-294, 2010.

ZIVIANI, N. A quarta revolução tecnológica. Computação cognitiva e a humanização das máquinas. **Fonte**, [s. l.], ano 14, n. 17, jul. 2017. Disponível em: <https://homepages.dcc.ufmg.br/~nivio/papers/a-quarta-revolucao-industrial-fonte-julho2017.pdf>. Acesso em: 10 maio 2026.

Recebido em 15/03/2026

Versão corrigida recebida em 11/05/2026

Aceito em 12/05/2026

Publicado online em 20/05/2026