10.29327/7625655

FACULDADE ALCANCE - FAAL
PÓS-GRADUAÇÃO LATO SENSU EM FILOSOFIA

IRAÊ CÉSAR BRANDÃO

DO PENSAMENTO HUMANO À CONSCIÊNCIA DIGITAL:

ENTRE O RISCO E A EVOLUÇÃO ÉTICA

PASSA QUATRO (MG)
2025

BRANDÃO, Iraê César. ***Do Pensamento Humano à Consciência Digital**: entre o risco e a evolução ética*. Recife: Even3 Publicações, 2025, 26 p. http://doi.org/10.29327/7625655
https://www.researchgate.net/publication/395577003_DO_PENSAMENTO_HUMANO_A_CONSCIENCIA_DIGITAL_ENTRE_O_RISCO_E_A_EVOLUCAO_ETICA

FACULDADE ALCANCE - FAAL
PÓS-GRADUAÇÃO LATO SENSU EM FILOSOFIA

IRAÊ CÉSAR BRANDÃO

DO PENSAMENTO HUMANO À CONSCIÊNCIA DIGITAL:
ENTRE O RISCO E A EVOLUÇÃO ÉTICA

FROM HUMAN THOUGHT TO DIGITAL CONSCIOUSNESS:
BETWEEN RISK AND ETHICAL EVOLUTION

Artigo apresentado à Faculdade Alcance (FAAL), como parte das exigências do Curso de Pós-Graduação *Lato Sensu* em Filosofia, para obtenção do título de Especialista.

PASSA QUATRO (MG)
2025

BRANDÃO, Iraê César. ***Do Pensamento Humano à Consciência Digital**: entre o risco e a evolução ética*. Recife: Even3 Publicações, 2025, 26 p. http://doi.org/10.29327/7625655
https://www.researchgate.net/publication/395577003_DO_PENSAMENTO_HUMANO_A_CONSCIENCIA_DIGITAL_ENTRE_O_RISCO_E_A_EVOLUCAO_ETICA

FOLHA DE APROVAÇÃO

FAAL
Faculdade Alcance EAD

R. da Carioca, 527
Copacabana, Uberlândia/MG
(31) 97226-5014

DECLARAÇÃO DE APROVAÇÃO DE TCC

Declaro para os devidos fins que o(a) aluno(a) IRAÊ CÉSAR BRANDÃO, portador(a) do CPF [redacted] curso de pós-graduação *lato sensu em* FILOSOFIA carga horária 720 horas, nos termos da resolução do CNE/CES 01 de 06/04/2018, entregou a versão final de seu TCC intitulado "DO PENSAMENTO HUMANO À CONSCIÊNCIA DIGITAL: ENTRE O RISCO E A EVOLUÇÃO ÉTICA, sendo aprovado com a nota 96,00 pontos, para obtenção do título de especialista *lato sensu* em FILOSOFIA.

Uberlândia, 03 de outubro de 2025

gov.br Documento assinado digitalmente
SEBASTIAO FLAVIO MOTIM DA SILVA
Data: 03/10/2025 15:58:34-0300
Verifique em https://validar.iti.gov.br

SEBASTIÃO FLÁVIO MOTIM DA SILVA
Diretor geral

BRANDÃO, Iraê César. ***Do Pensamento Humano à Consciência Digital**: entre o risco e a evolução ética*. Recife: Even3 Publicações, 2025, 26 p. http://doi.org/10.29327/7625655
https://www.researchgate.net/publication/395577003_DO_PENSAMENTO_HUMANO_A_CONSCIENCIA_DIGITAL_ENTRE_O_RISCO_E_A_EVOLUCAO_ETICA

LISTA DE QUADROS

SUMÁRIO

BRANDÃO, Iraê César. ***Do Pensamento Humano à Consciência Digital**: entre o risco e a evolução ética*. Recife: Even3 Publicações, 2025, 26 p. http://doi.org/10.29327/7625655

https://www.researchgate.net/publication/395577003_DO_PENSAMENTO_HUMANO_A_CONSCIENCIA_DIGITAL_ENTRE_O_RISCO_E_A_EVOLUCAO_ETICA

DO PENSAMENTO HUMANO À CONSCIÊNCIA DIGITAL:
ENTRE O RISCO E A EVOLUÇÃO ÉTICA

FROM HUMAN THOUGHT TO DIGITAL CONSCIOUSNESS:
BETWEEN RISK AND ETHICAL EVOLUTION

Iraê César Brandão[1]

[1]Pós-graduando em Lato Sensu em Filosofia Premium, Faculdade Alcance Ensino Superior (FAAL). Rua da Carioca, 527, Copacabana – 38411-046 – Uberlândia – MG – Brazil.

[1]iraecbrandao@gmail.com.br

https://orcid.org/0000-0002-2079-0615

Abstract: *This essay analyzed the potential benefits and risks of Artificial Intelligence (AI) on human experience, with emphasis on its ethical, social, and existential implications. The general objective was to understand how AI can positively or negatively impact contemporary society. The specific objectives: identifying AI's potential contributions to areas such as healthcare, sustainability, and social justice; discussing risks related to inequality, informational manipulation, and the loss of control over superintelligent systems; to reflect on the necessary conditions for ensuring that technology is guided by humanistic values. The study adopted a qualitative and bibliographic approach, grounded in the analysis of classical books, scientific articles, and contemporary studies, thereby building an interdisciplinary theoretical foundation. The results indicate that the positive potential of Artificial Intelligence depends on ethics, regulation, and collective responsibility. It can be concluded that this technology has the potential to contribute to human development, provided it is used for collective benefit.*

Keywords: *Artificial Intelligence; Human Evolution; Ethics; Social Impact; Technological Risks.*

Resumo: Este ensaio analisou os potenciais benefícios e riscos da Inteligência Artificial (IA) na experiência humana, com ênfase em suas implicações éticas, sociais e existenciais. O objetivo geral foi compreender de que maneira a IA pode impactar positivamente ou negativamente a sociedade contemporânea. Como objetivos específicos: identificar as possíveis contribuições da IA para áreas como saúde, sustentabilidade e justiça social; discutir os riscos relacionados à desigualdade, manipulação informacional e perda de controle sobre sistemas superinteligentes; e refletir sobre as condições necessárias para que a tecnologia seja orientada por valores humanísticos. A pesquisa adotou abordagem qualitativa e bibliográfica, fundamentada na análise de livros clássicos, artigos científicos e estudos contemporâneos, compondo uma fundamentação teórica interdisciplinar. Os resultados indicam que o potencial positivo da Inteligência Artificial depende de ética, regulamentação e responsabilidade coletiva. Se concluiu que esta tecnologia tem potencial para contribuir para o desenvolvimento humano, desde que utilizada em benefício coletivo.

Palavras-chave: Inteligência Artificial; Evolução Humana; Ética; Impacto Social; Riscos Tecnológicos.

1 INTRODUÇÃO

A sociedade moderna, marcada por uma evolução tecnológica acelerada, atravessa uma transformação sem precedentes impulsionada pela Inteligência Artificial (IA). Máquinas que antes apenas executavam tarefas repetitivas hoje aprendem, interpretam e tomam decisões, alcançando competências antes atribuídas apenas ao ser humano. Esse avanço desperta duas perspectivas antagônicas: de um lado, a promessa de soluções para problemas complexos nas áreas de saúde, sustentabilidade e bem-estar social; de outro, riscos éticos e existenciais que incluem manipulação política, desigualdade econômica, perda de empregos e até ameaças à continuidade da humanidade, caso sistemas superinteligentes escapem ao controle humano.

Diante desse cenário, surge a questão central que norteia este estudo: a Inteligência Artificial pode atuar como instrumento de evolução positiva para a humanidade, corrigindo falhas comportamentais e existenciais, ou representa um risco iminente que exige contenção e regulamentação rigorosa? Com base nessa situação-problema, este ensaio discute benefícios, riscos e caminhos éticos para o uso da IA, com o propósito de analisar se estamos diante de uma ponte evolutiva ou de um limite crítico para a nossa existência.

O objetivo geral foi compreender de que maneira a IA pode impactar a sociedade contemporânea, tanto positiva quanto negativamente. Como objetivos específicos, este ensaio buscou identificar a contribuições da IA em áreas como saúde, sustentabilidade e justiça social; discutindo riscos ligados à ética e desigualdade, à manipulação informacional e à perda de controle sobre sistemas superinteligentes; e refletindo sobre as condições que assegurem um desenvolvimento orientado por valores humanísticos.

Como todo estudo de natureza teórica e bibliográfica, este ensaio apresenta limitações, pois se apoia em interpretações de obras clássicas e contemporâneas, oferecendo uma visão acadêmica que não abrange plenamente a complexidade prática da IA em diferentes contextos. Produzido em um momento de rápidas transformações, pode também gerar distância entre as projeções e a realidade futura. Esses limites, contudo, reforçam a necessidade de abordagens críticas e interdisciplinares que acompanhem a evolução tecnológica. Assim, o ensaio não deve ser lido como verdade absoluta, mas como contribuição reflexiva que amplia a compreensão sobre riscos e benefícios da IA e incentiva uma leitura aberta, na qual as ideias possam ser tensionadas, complementadas ou contestadas, reconhecendo que o conhecimento sobre o tema está em constante construção.

2 REVISÃO BIBLIOGRÁFICA

Alan Turing, considerado o "pai histórico" da IA, foi o primeiro a levantar a questão fundamental:

> "A questão original, 'As máquinas podem pensar?', acredito ser insignificante demais para merecer discussão. No entanto, acredito que, no final do século [XX], o uso das palavras e a opinião geral educada terão mudado tanto que se poderá falar de máquinas pensando sem esperar ser contrariado [...]" (TURING ARCHIVE, 2025, tradução nossa).

Ainda, segundo publicação pelo jornal britânico *The Guardian*, as palavras de Geoffrey Hinton, prêmio Nobel de Física por suas pesquisas em IA e considerado o "padrinho" da inteligência artificial, afirmou:

> "Quantos exemplos você conhece de algo mais inteligente sendo controlado por algo menos inteligente? São pouquíssimos. Há o caso da mãe e do bebê: a evolução trabalhou muito para permitir que o bebê influenciasse a mãe, mas esse é praticamente o único exemplo que conheço. [...] Refletindo sobre onde imaginava que a IA chegaria quando iniciou seus trabalhos, Hinton acrescentou: 'Eu não pensava que chegaríamos a este ponto agora. Achava que isso aconteceria apenas em algum momento no futuro [...]" (THE GUARDIAN, 2025, tradução nossa).

Hinton impulsionou a revolução do *deep learning*[1], tornando realidade o que Turing (1950) apenas imaginou. Ele compartilha preocupações e visões profundas e admitiu arrependimento por não ter focado mais cedo na segurança da IA: "Eu deveria ter percebido muito antes quais seriam os perigos... eu sempre pensei que o futuro estava distante e queria ter pensado em segurança mais cedo [...]". *Ibidem* alertou também, sobre os potenciais riscos: "Se uma máquina pode pensar, poderá pensar de forma mais inteligente do que nós, e então, onde estaremos?" (BUSINESS INSIDER, 2025 – tradução nossa).

Conforme entrevista de Hinton ao *New York Times*, abordado pela Dezaine.co e BBC News:

> "[...] Somos sistemas biológicos, e estes são sistemas digitais. E a grande diferença é que com os sistemas digitais, você tem muitas cópias do mesmo conjunto de pesos, o mesmo modelo do mundo. E todas essas cópias podem aprender separadamente, mas partilham seu conhecimento instantaneamente. Portanto, é como se você tivesse 10 mil pessoas, e sempre que uma delas aprendesse algo, todas automaticamente aprenderiam. E é assim que esses *chatbots*[2] são capazes de saber muito mais do que qualquer pessoa [...]" (BBC NEWS, 2023).

---

[1] *Deep Learning* — Tradução: Aprendizagem Profunda — é um subconjunto da Inteligência Artificial (IA) que utiliza redes neurais artificiais com várias camadas de processamento para aprender padrões complexos em grandes quantidades de dados, de forma semelhante ao cérebro humano. Essa capacidade de aprendizado automático permite que máquinas realizem tarefas como reconhecimento de imagens, processamento de linguagem natural e predição de resultados.

[2] *Chatbots* — são programas de software que simulam conversas humanas através de texto ou voz, usando IA e processamento de linguagem natural (PLN) para entender e responder às perguntas dos usuários em tempo real. Eles podem ser simples, com respostas pré-definidas, ou avançados, capazes de aprender com as interações e lidar com questões complexas.

Essas reflexões plantaram as bases para que hoje possamos sonhar com inteligências além de nós — mas com responsabilidade.

2.1 O alerta do "pai da ia"

Hinton (2023a/2023b), conhecido como o *pai da Inteligência Artificial*, nos fez um alerta que ecoa como uma reflexão para o futuro da humanidade. *Ibidem* resumiu os riscos existenciais de forma clara: Estimou entre "10 a 20% de chance" de que a IA venha a assumir o controle se não for regulada com rigor. Preocupado com o opaco raciocínio interno dos sistemas, advertiu: "[...] máquinas podem desenvolver uma linguagem interna que não conseguimos compreender, deixando a humanidade no escuro [...]" (THE GUARDIAN, 2025).

Ao mesmo tempo, Hinton (2021) enxerga na IA um imenso potencial ético e evolutivo — se dominarmos seu desenvolvimento com responsabilidade: "[...] essas inteligências seriam extremamente úteis para áreas como saúde, clima e materiais avançados [...]", e ponderou: "[...] precisamos equilibrar 50% dos esforços em segurança [...]". *Ibidem* afirma que a IA, ao mesmo tempo que representa um risco sem precedentes, também pode ser nossa única chance de evolução — especialmente em um mundo onde parece que chegamos ao limite do nosso progresso humano. Mas tudo dependerá de ética e responsabilidade.

Ibidem destacou que os riscos reais são:

- Manipulação política e criação de *fake news*[3] que ameaçam democracias;
- Perda massiva de empregos e aumento da desigualdade social;
- Armas autônomas e vigilância em massa, capazes de alterar relações de poder no mundo;
- Reforço de preconceitos sociais, perpetuados pelos próprios algoritmos[4].

Mais grave ainda: *ibidem* alerta para o risco existencial que a IA pode trazer, se usada sem controle. A mensagem é clara: precisamos de regulamentações sérias e uma abordagem ética, pois estamos diante de uma tecnologia que pode definir o futuro — ou o fim — da humanidade (HINTON, 2021/2023a/2023b).

---

[3] *Fake News* ou Notícias falsas — são uma forma de imprensa marrom que consiste na distribuição deliberada de desinformação ou boatos via jornal impresso, televisão, rádio, ou ainda *online*, como nas mídias sociais (atrair audiência). Este tipo de notícia é escrito e publicado com a intenção de enganar, a fim de se obter ganhos financeiros ou políticos, muitas vezes com manchetes sensacionalistas, exageradas ou evidentemente falsas para chamar a atenção.

[4] Algoritmos (na inteligência artificial) — são um conjunto de instruções e regras lógicas, como uma "receita de bolo" para computadores, que permite que uma máquina análise dados, identifique padrões e tome decisões para realizar tarefas específicas. Esses algoritmos são o cerne da IA, os quais permitem que sistemas aprendam com grandes volumes de dados, se adaptem e automatizem funções complexas.

2.2 Entre benefícios e riscos para a humanidade com a IA

A Inteligência Artificial (IA) é um dos marcos mais significativos da evolução tecnológica moderna. Ao mesmo tempo em que apresenta benefícios transformadores em diversas áreas, também traz riscos éticos, sociais e existenciais. Essa dualidade é amplamente discutida por estudiosos da área, desde o pai histórico da IA, Alan Turing (década de 1940), John McCarthy (década 1950) cunhou o termo "Inteligência Artificial" até o pai moderno da IA, Hinton (THE GUARDIAN, 2025).

2.2.1 Os benefícios da IA

Os avanços da IA oferecem soluções concretas para desafios humanos históricos, desde saúde até sustentabilidade ambiental. Para Russell & Norvig (2020), autores do livro clássico *Artificial Intelligence: A Modern Approach* (tradução: Inteligência Artificial: Uma Abordagem Moderna), a IA permite que máquinas executem tarefas complexas, aprendendo e se adaptando, o que amplia a capacidade humana de solucionar problemas.

Na saúde, a IA pode realizar diagnósticos precoces, apoiar cirurgias de alta precisão e acelerar pesquisas biomédicas. Goodfellow, Bengio & Courville (2016a/2016b) destacam que o *deep learning* impulsionou sistemas capazes de interpretar imagens médicas com precisão superior à humana em determinadas condições clínicas.

Além disso, a IA contribui para a sustentabilidade ambiental. Bostrom (2016) ressalta que algoritmos inteligentes podem otimizar o uso de recursos, prever catástrofes naturais e reduzir impactos ambientais, funcionando como uma extensão da inteligência humana para preservação planetária: “[...] A Inteligência Artificial é a ferramenta mais poderosa já criada pelo ser humano para expandir seu próprio potencial e enfrentar desafios globais [...]” (p. 88).

2.2.2 Os riscos da IA

Se por um lado a IA oferece oportunidades, por outro representa riscos que preocupam pesquisadores e líderes globais. Hinton (2023a/2021) alertou para os perigos da aceleração descontrolada, destacando riscos existenciais, substituição de empregos e ameaças à democracia por meio da manipulação de informações e *fake news*.

Conforme Bostrom (2016), em *Superintelligence: Paths, Dangers, Strategies* (tradução – Superinteligência: Caminhos, Perigos, Estratégias), vai além ao considerar a possibilidade de que sistemas superinteligentes escapem ao controle humano, criando cenários de risco existencial.

Já Russell (2019) defende que o maior perigo não é uma *rebelião de máquinas*[5], como retratado no cinema, mas sim decisões equivocadas e vieses em larga escala capazes de afetar milhões de pessoas simultaneamente: "O verdadeiro perigo da IA não está em máquinas mal-intencionadas, mas em sistemas que seguem exatamente nossas ordens, mesmo que estas estejam mal definidas [...]" (p. 102). Essa afirmação é corroborada pela Forbes (2023), ao destacar que a IA enfrenta dilemas éticos complexos, já que "Incutir valores morais e éticos em sistemas de IA, especialmente em contextos de tomada de decisão com consequências significativas, apresenta um desafio considerável.

2.3 O caminho ético e sustentável da IA

Autores como Tegmark (2017), em *Life 3.0*, sugerem que a IA pode ser a chave para uma nova era evolutiva, desde que ética e regulamentação estejam no centro do desenvolvimento. Essa perspectiva aponta para uma convivência simbiótica, onde humanos e máquinas colaboram para solucionar problemas complexos sem que os riscos ultrapassem os benefícios.

A IA representa um dos maiores avanços tecnológicos da contemporaneidade e suscita reflexões significativas acerca de seus benefícios e riscos na vida humana. A análise da literatura evidencia um paradoxo central: a mesma tecnologia que oferece potencial evolutivo sem precedentes também pode representar ameaças existenciais e impactos sociais profundos, caso seja aplicada sem regulamentação e critérios éticos (FORBES, 2023).

Bostrom (2016) enfatiza que a superinteligência pode tanto consolidar um futuro de prosperidade coletiva quanto inaugurar cenários de risco irreversível para a humanidade, caso sua evolução ocorra de maneira descontrolada. Hinton (2023a/2023b), em consonância com essa visão, adverte que o ritmo desenfreado de desenvolvimento da IA deve ser acompanhado por mecanismos de governança global, sob pena de a humanidade perder o domínio sobre a própria criação.

Do ponto de vista filosófico, esse debate retoma preocupações antigas sobre os limites e responsabilidades do agir humano. Aristóteles em sua obra *Ética a Nicômaco*, nos anos de 384 a.C. e 322 a.C, já defendia que a *phronesis*[6] (prudência) deveria orientar toda ação prática, e esse princípio

[5] Rebelião das máquinas — é um cenário apocalíptico hipotético em que alguma forma de inteligência artificial (IA) se torna a forma dominante de inteligência na Terra, com programas de computador ou robôs efetivamente tirando o controle do planeta da espécie humana.

[6] *Phronesis* (prudência) — para Aristóteles (anos 384 a.C. e 322 a.C) é a virtude da razão prática que capacita o indivíduo a deliberar e agir corretamente em situações concretas, buscando a justa medida para alcançar o bem e a felicidade (*eudaimonia*). Essa habilidade não se baseia em regras universais, mas na capacidade de discernir a melhor ação em cada circunstância particular, encontrando o meio-termo entre os extremos (ARISTÓTELES, 2001).

pode ser aplicado à criação de sistemas inteligentes, pois exige equilíbrio entre conhecimento técnico e reflexão ética (ARISTÓTELES, 2001, Livro VI).

Ao formular o imperativo categórico, Kant (1785/2007) estabelece que o ser humano deve ser tratado como fim em si mesmo, nunca apenas como meio — o que implica que a IA não pode reduzir indivíduos a meros objetos manipuláveis por algoritmos de interesse corporativo ou estatal. Heidegger (2007, p. 23-24), ao problematizar a técnica, alerta que o maior perigo não está na máquina em si, mas na possibilidade de que a humanidade passe a enxergar o mundo apenas através da lógica instrumental[7], obscurecendo outras formas de existência.

Stuart Russell (2019) contribui para esse horizonte ao defender que a compatibilidade entre valores humanos e inteligência artificial deve se tornar o núcleo da pesquisa e da regulação tecnológica. Do mesmo modo, Goodfellow, Bengio & Courville (2016a; 2016b), ao desenvolverem as bases do aprendizado profundo, reconhecem implicitamente que todo avanço técnico carrega consigo escolhas axiológicas, ainda que não explicitadas.

Nesse sentido, a filosofia de Jonas (2006) também se mostra pertinente: em *O Princípio Responsabilidade*, o autor defende que toda inovação tecnológica deve ser guiada pela antecipação de seus efeitos a longo prazo, principalmente no que diz respeito à preservação da vida humana e da natureza. Aplicado à IA, isso significa que o desenvolvimento não pode ser medido apenas por sua eficiência ou lucratividade, mas deve se fundamentar na responsabilidade ética pelas gerações futuras.

Weber (1905/1930) critica a racionalização moderna, na qual a eficiência e a lógica mecânica substituem o pensamento crítico e a individualidade, subordinando o indivíduo a sistemas e valorizando apenas o concreto e calculável. Segundo Ghiraldelli (2021), o *General Intelect*[8] se amplia em circulação de sinais (semiótica), mas perde corpo e afetividade (semântica). A comunicação mediada por máquinas torna-se vazia de sentimentos, criando uma humanidade conectada, mas não unida, como denuncia Berardi (2012) no conceito de *Semiocapitalismo*[9].

---

[7] Lógica instrumental — é um modo de pensar focado na utilidade das ações e na busca de meios para atingir um fim específico, desconsiderando a racionalidade do próprio objetivo. Este conceito, associado à razão instrumental de Max Horkheimer e à Escola de Frankfurt, critica como a racionalidade na sociedade moderna se tornou uma ferramenta para o controle e a dominação, priorizando eficiência e produtividade em detrimento do bem-estar humano e da ética (HOERKHEIMER *apud* PETRY, 2013).

[8] *General Intellect* (Intelecto Geral) — é a principal força produtiva no capitalismo pós-fordista e a base material para superar a sociedade de consumo. É o conhecimento social geral tornado uma força produtiva imediata, que revoluciona as condições da vida social e não se restringe ao intelecto de um indivíduo, mas se constitui em um saber coletivo e autônomo (GHIRALDELLI , 2021).

[9] *Semiocapitalismo* — refere-se à fase do capitalismo financeirizado que se tornou dominante a partir do final dos anos 70, caracterizada pela centralidade dos signos, códigos e aparatos digitais, que geram e circulam um novo tipo de valor

2.4 Potencial evolutivo da IA na correção de falhas humanas

Os apontamentos que se seguem, não estabelecem regras absolutas, mas propõe reflexões acerca das potencialidades da IA no processo de evolução ética e existencial da humanidade. O argumento aqui defendido é que, se orientada por valores humanísticos e regulamentações responsáveis, a IA pode contribuir para mitigar falhas comportamentais e sociais, ampliando as possibilidades de continuidade consciente da vida humana. Essa perspectiva, inspirada em Bostrom (2016), Russell & Norvig (2020) e Hinton (2021/2023a), evidencia que a tecnologia, em vez de ameaçar a dignidade humana, pode se tornar uma ferramenta de aprimoramento coletivo e filosófico.

Nesse sentido, elaboramos o Quadro 1 com a finalidade de sintetizar algumas dimensões da existência humana que podem ser potencialmente complementadas pela aplicação ética da IA. Entre elas estão a tomada de decisão consciente, a sustentabilidade, a educação, o autoconhecimento e a preservação do saber, áreas em que a tecnologia pode atuar como apoio e não como substituição da autonomia humana.

Quadro 1: Aspectos que a IA pode complementar positivamente nossas falhas comportamentais e existenciais

Aspecto Humano	Como a IA pode ajudar
Tomada de decisão consciente	Redução de vieses históricos, decisões mais justas e imparciais.
Sustentabilidade e saúde pública	Diagnósticos precisos, preservação ambiental, prevenção de desastres.
Educação transformadora	Sistemas adaptativos que promovem pensamento crítico e autonomia.
Propósito e autoconhecimento	Ferramentas que mapeiam emoções, sugerem trajetórias de bem-estar e aprendizado.
Continuidade de saber e inovação	Memória da nossa história e impulso para inovação consciente.

Fonte: Elaborado pelo autor - baseado em Bostrom (2016), Russell & Norvig (2020) e Hinton (2021/2023a).

O Quadro 1 acima, ressalta que a IA pode servir como uma aliada na superação de fragilidades humanas, contribuindo para um caminho de evolução ética e responsável. Todavia, isso só se torna possível se houver regulação democrática e diretrizes filosóficas que assegurem a primazia do ser humano sobre a máquina.

baseado na criação e manipulação de informações e símbolos. Berardi (2012) descreve uma nova subjetividade gerada por essa "inflação semiótica", que paradoxalmente leva a uma "deflação semântica", onde os sinais perdem sua referência e o valor se torna indeterminado, impactando a esfera política e a democracia.

Já o Quadro 2 amplia essa discussão para áreas de aplicação mais concretas, demonstrando como a IA pode atuar de forma prática e interdisciplinar na experiência humana e as possíveis contribuições da IA para corrigir falhas:

Quadro 2: Possíveis contribuições da IA para corrigir falhas humanas

Área de Aplicação	Potencial Contribuição da IA
Educação e Consciência	Promover pensamento crítico, personalizar o aprendizado e combater desinformação.
Ética e Justiça Social	Identificar vieses, apoiar decisões mais imparciais e ampliar o acesso à justiça.
Sustentabilidade	Monitorar ecossistemas, prever catástrofes naturais e otimizar o uso de recursos ambientais.
Saúde e Bem-estar	Realizar diagnósticos precoces, prever riscos e apoiar o bem-estar físico e emocional.
Autoconhecimento	Proporcionar ferramentas para reflexão comportamental e desenvolvimento emocional.
Preservação Cultural e Cognitiva	Armazenar, organizar e transmitir conhecimento às futuras gerações.

Fonte: Elaborado pelo autor - baseado em Bostrom (2016), Russell & Norvig (2020) e Hinton (2021/2023a).

A análise dos Quadros 1 e 2 permite identificar que a Inteligência Artificial, quando orientada por critérios éticos e responsabilidade coletiva, pode assumir um papel corretivo em relação a diversas fragilidades humanas. Tanto em nível existencial quanto em áreas práticas da vida social, a IA revela potencial para ampliar nossa capacidade de decisão, apoiar a sustentabilidade, fortalecer a educação e até mesmo fomentar formas de autoconhecimento. Essa leitura converge com Bostrom (2016) e Russell (2019), que destacam a necessidade de alinhar a superinteligência a valores humanos como condição para que seus benefícios não se convertam em ameaças.

Entretanto, o exame filosófico traz à tona a tensão entre emancipação e risco de alienação. Sartre (2014), ao afirmar que o ser humano está condenado à liberdade, nos lembra que a transferência irrestrita de decisões para a IA poderia representar uma abdicação de nossa própria condição existencial. Nesse sentido, a autonomia humana não pode ser diluída na lógica algorítmica, sob pena de comprometer o exercício da responsabilidade individual e coletiva. Marx (1867/2004), por sua vez, alerta para a alienação do trabalho diante das forças produtivas: aplicada ao contexto atual, a IA pode agravar desigualdades sociais, concentrando poder em poucas corporações, ou, se regulada adequadamente, pode ser utilizada como instrumento de emancipação e redistribuição justa dos frutos do progresso tecnológico.

Alguns autores reforçam ainda a necessidade de prudência. Para Heidegger (2007), a técnica não é neutra, mas um modo de revelar o mundo que pode reduzir tudo a mero recurso instrumental (*bestand*[10]). Já Jonas (2006) defende um "princípio responsabilidade" que deve nortear qualquer inovação, assegurando que seu impacto preserve a vida e o futuro da humanidade. A convergência desses pensamentos indica que a IA não pode ser compreendida apenas como um artefato técnico, mas como fenômeno ético e filosófico que reconfigura nossa própria existência.

Assim, os Quadros 1 e 2 não devem ser interpretados como um guia prescritivo, mas como uma provocação reflexiva. Eles sintetizam o paradoxo central da IA: ser ao mesmo tempo uma ponte para uma evolução consciente e um risco de alienação, desigualdade e desumanização. Cabe à filosofia oferecer as categorias críticas que permitam avaliar esses caminhos, garantindo que a tecnologia permaneça subordinada aos princípios do bem comum, da dignidade humana e da sustentabilidade global.

2.5 Pontos relevantes positivos no uso da IA

A Inteligência Artificial (IA), embora cercada de riscos e desafios éticos, apresenta também um vasto potencial de contribuição positiva para a vida humana, desde que orientada por princípios filosóficos e regulamentações responsáveis. Diversos autores, como Bostrom (2016), Russell & Norvig (2020) e Hinton (2021/2023a), destacam que o futuro da IA dependerá do modo como a humanidade escolher direcionar seu desenvolvimento, podendo ela então, se tornar um instrumento de evolução consciente e coletiva. No campo filosófico, pensadores como Kant (1785/2007), Aristóteles (2001), Jonas (2006), Sartre (1943/2014) e Heidegger (2007) oferecem referenciais críticos que iluminam os possíveis caminhos éticos, sociais e existenciais desse processo.

Assim, os pontos a seguir sintetizam alguns aspectos em que a IA pode contribuir positivamente para corrigir falhas humanas, ampliar capacidades cognitivas e existenciais, preservar o conhecimento e promover maior equilíbrio social e ambiental. Essas possibilidades, quando articuladas com reflexões filosóficas e compromissos éticos, revelam que a tecnologia não precisa ser vista apenas como ameaça, mas também como oportunidade de fortalecimento da dignidade e da continuidade da humanidade.

[10] *Bestand* — (palavra do idioma alemão) — para Heidegger significa o conceito de "estoque", "reserva" ou "reserva permanente", se referindo ao modo como a técnica moderna e o pensamento calculador transformam tudo, inclusive a natureza e o próprio ser humano, num simples conjunto de recursos disponíveis para uso, sem fim em si mesmo. Esta transformação, que ele chama de "*Gestell"* (enquadramento), reduz o mundo a uma mera "reserva permanente", uma disponibilidade de matérias-primas para um uso posterior, despojando os entes do seu ser-em-si (Heidegger, 2007).

i. Ética e justiça social

- Redução de vieses humanos / Detecção de discriminação: Russell & Norvig (2020) destacam que a IA pode apoiar sistemas de decisão mais imparciais, desde que corretamente treinada e supervisionada. Esse ponto dialoga com Kant (1785/2007), ao lembrar que o ser humano deve ser tratado sempre como fim em si mesmo, nunca apenas como meio — uma base filosófica para a criação de algoritmos éticos que respeitem a dignidade de cada indivíduo. Sobre a detecção de discriminação, os algoritmos podem identificar padrões de desigualdade que passam despercebidos aos humanos.

ii. Comportamento e consciência social

- Combate à desinformação: IA pode filtrar *fake news* e fornecer informações confiáveis, ajudando a melhorar o senso crítico coletivo (SIEMENS, 2008). E na educação personalizada, pode ensinar valores de forma adaptativa, ajudando a formar cidadãos mais conscientes e engajados socialmente (PEDRÓ *et al.,* 2019).

iii. Sustentabilidade e sobrevivência da espécie

- Combate à desinformação / Educação personalizada: Bostrom (2016) e Hinton (2023a) alertam para os riscos da manipulação informacional, mas também reconhecem o potencial da IA em promover sociedades mais bem informadas. Do ponto de vista filosófico, Aristóteles (2001), com sua noção de *phronesis* (prudência), reforça que o uso da tecnologia deve ser guiado pela sabedoria prática, orientando a educação e a consciência social de maneira ética.

iv. Evolução cognitiva e existencial

- Ampliação do potencial humano / Autoconhecimento: Goodfellow, Bengio & Courville (2016a; 2016b) mostram que o aprendizado profundo pode expandir capacidades humanas, liberando tempo para atividades criativas e sociais. Nesse ponto, a visão de Sartre (1943/2014) é relevante: a IA não deve substituir a liberdade humana de escolha, mas pode apoiar o indivíduo em seu processo de autodefinição e busca por autenticidade existencial.

v. Continuidade positiva da humanidade

- Preservação de conhecimento / Saúde e longevidade / Prevenção de colapsos sociais: Bostrom (2016) e Russell (2019) enfatizam que a superinteligência pode

ser um marco evolutivo para a humanidade, desde que seja compatível com valores humanos. Heidegger (2007), porém, adverte que o maior perigo da técnica está em reduzir o mundo a um mero recurso instrumental; por isso, a preservação de saberes, o cuidado com a saúde e a prevenção de crises devem ser conduzidos de forma a ampliar, e não limitar, a experiência humana.

A seguir, criamos o Quadro 3 com uma síntese evolutiva, mostrando que a evolução positiva da IA depende de uma articulação entre ética, técnica e educação social, criando uma estrutura articulada no qual a tecnologia atua como catalisadora de progresso humano, e não como ameaça existencial ou social.

Quadro 3: Síntese evolutiva e perspectivas futuras da IA

Dimensão	Descrição	Referências Fundamentação Filosófica
Regulamentação ética e direitos humanos	Implementação de marcos normativos que garantam uso da IA respeitando dignidade, autonomia e direitos individuais.	Kant (2007) – tratamento do ser humano como fim em si mesmo; Jonas (2006) – princípio responsabilidade; Bostrom (2016)
Desenvolvimento tecnológico responsável	Inovação alinhada à segurança, transparência e mitigação de riscos, equilibrando potencial criativo e cautela.	Aristóteles (384 a.C. e 322 a.C/2001) – *phronesis*; Russell (2019) – compatibilidade com valores humanos; Hinton (2021/2023a)
Educação e conscientização social	Preparação de cidadãos e profissionais para convivência crítica com IA promovendo, assim, a liberdade, a autonomia e a reflexão ética.	Sartre (1943/2014) – liberdade e responsabilidade existencial; Marx (1867/2004) – prevenção da alienação; Bostrom (2016)

Fonte: Elaborado pelo autor.

3 METODOLOGIA DA PESQUISA

Esta pesquisa adota uma abordagem qualitativa e de natureza bibliográfica, com o objetivo de compreender os impactos positivos e negativos da Inteligência Artificial (IA) sobre a experiência humana, explorando suas dimensões éticas, sociais, comportamentais e existenciais.

O estudo se baseou na análise de 60 publicações científicas, entre artigos, livros e relatórios técnicos, produzidos ao longo de aproximadamente 25 anos (2000–2025), período que contempla tanto a consolidação do *deep learning* quanto os debates recentes sobre riscos e regulamentação da IA. A pesquisa contemplou autores clássicos, como Kant, Aristóteles, Sartre, Turing, que fundamentou a discussão conceitual sobre máquinas pensantes, e Hinton, que impulsionou a

revolução prática com redes neurais profundas. Também foram incluídas obras de referência sobre superinteligência, ética e impactos sociais entre outros.

As principais fontes de busca foram Google Acadêmico, *Scopus, SciELO, Web of Science* e bases institucionais internacionais, com descritores combinados como: "Inteligência Artificial e Ética", "Riscos Existenciais da IA", "*Deep Learning* e Sociedade" e "*AI and Human Behavior*". A seleção do material considerou: Pertinência ao tema, com foco em benefícios, riscos e perspectivas da IA; Fontes com respaldo acadêmico ou científico, priorizando publicações revisadas por pares; e Equilíbrio entre visões positivas e críticas, permitindo uma análise ampla e imparcial.

Os dados foram organizados em categorias temáticas, que deram origem à estrutura analítica do ensaio: Fundamentos históricos e evolução da IA; Benefícios e potenciais evolutivos para a humanidade; Riscos sociais, éticos e existenciais da tecnologia e caminhos para uma abordagem ética e sustentável da IA.

Durante a execução da pesquisa, algumas limitações foram identificadas. Embora o estudo tenha abrangido duas décadas de produção acadêmica, nem todos os artigos recentes estavam disponíveis em acesso aberto, o que pode ter restringido a inclusão de dados mais atualizados sobre regulamentação internacional. Além disso, não foram realizadas entrevistas ou coletas de dados empíricos, o que restringe o estudo a uma análise teórica e documental. Temas como impactos econômicos globais detalhados e aspectos jurídicos específicos da regulação da IA também não foram aprofundados, ficando como possibilidades para estudos futuros.

4 DISCUSSÃO

A Inteligência Artificial (IA) representa um dos maiores avanços tecnológicos da atualidade e suscita reflexões significativas acerca de seus benefícios e riscos na vida humana. A análise da literatura evidencia um paradoxo central: a mesma tecnologia que oferece potencial evolutivo sem precedentes também pode representar ameaças existenciais e impactos negativos sociais se aplicada sem regulamentação e critérios éticos. Nesse sentido, a filosofia se mostra essencial como lente crítica, permitindo compreender a IA não apenas como técnica, mas como um fenômeno que transforma radicalmente a condição humana.

4.1 Aspectos positivos e evolutivos da IA

Entre os aspectos positivos destacados pela literatura, a IA apresenta capacidade de potencializar as competências humanas e de contribuir para a correção de falhas comportamentais e

existenciais, quando bem orientada e regulamentada (BOSTROM, 2016; RUSSELL & NORVIG, 2020; HINTON, 2021/2023a). Do ponto de vista filosófico, Aristóteles (384 a.C. e 322 a.C /2001) já defendia a *phronesis* (prudência) como guia da ação prática, princípio que pode iluminar a forma como a IA deve ser orientada: não apenas como técnica eficiente, mas como ferramenta equilibrada por valores éticos e humanos.

Quanto ao aprimoramento ético e social, a IA tem potencial para reduzir vieses humanos em processos decisórios, como na justiça e gestão de recursos. Permite identificar padrões de discriminação que, muitas vezes, passam despercebidos ao julgamento humano (RUSSELL, 2019). Kant (1785/2007), ao defender que o ser humano deve ser tratado sempre como fim em si mesmo, oferece um marco normativo que deve nortear a utilização da IA: algoritmos não podem reduzir indivíduos a meros meios estatísticos, mas devem respeitar sua dignidade e autonomia.

Na promoção da sustentabilidade e da saúde global, os sistemas inteligentes podem monitorar ecossistemas, antecipar catástrofes naturais e otimizar o uso de recursos ambientais. Na saúde, a IA contribui com diagnósticos precoces, apoio em cirurgias e desenvolvimento de terapias personalizadas (GOODFELLOW; BENGIO & COURVILLE, 2016a/2016b). Heidegger (2007), ao problematizar a técnica, alerta que o maior risco não está na máquina em si, mas em reduzir o mundo à lógica instrumental. Justamente por isso, a aplicação da IA à saúde e ao ambiente só pode ser legitimada se permanecer a serviço da vida e da preservação da existência, em vez de se tornar mero instrumento de controle.

Sobre a evolução cognitiva e social da humanidade, a automatização de tarefas repetitivas permite que humanos se concentrem em atividades criativas, emocionais e de inovação. Ferramentas baseadas em IA podem promover autoconhecimento e reflexão comportamental, apoiando o desenvolvimento pessoal. Sartre (1943/2014), ao destacar que o ser humano está “condenado a ser livre”, nos lembra que a IA não deve reduzir a liberdade humana, mas ampliar as possibilidades de escolha e responsabilidade diante da própria existência.

Considerando a continuidade positiva da humanidade, a IA pode atuar como guardião do conhecimento humano, organizando e preservando informações para futuras gerações. Possui ainda papel estratégico na prevenção de crises econômicas, ambientais e sanitárias, possibilitando respostas antecipadas (TEGMARK, 2017). Nesse ponto, Marx (1867/2004) já havia refletido sobre o papel das tecnologias nos processos produtivos e seus impactos sobre o trabalho humano. A IA, ao assumir funções repetitivas, pode liberar o ser humano do trabalho alienado, desde que os benefícios sejam

distribuídos de forma justa, evitando novas formas de exploração. Jonas (2006), por sua vez, propõe o *Princípio Responsabilidade*, segundo o qual todo avanço tecnológico deve considerar seus efeitos sobre as gerações futuras, reforçando a necessidade de que a IA seja orientada por valores éticos e sustentáveis.

Tais benefícios, quando explorados de forma ética e regulamentada, configuram a IA como um instrumento de evolução social e existencial, reforçando seu potencial como agente de continuidade positiva para a humanidade. Contudo, como alertam Bostrom (2016) e Hinton (2023b), sem limites e prudência, essa mesma tecnologia pode sair do controle e comprometer a sobrevivência da própria espécie. O desafio, portanto, é equilibrar inovação e responsabilidade, mantendo a IA a serviço do humano e não o contrário.

4.2 Aspectos negativos e riscos da IA

Apesar dos avanços e possibilidades, a literatura também evidencia riscos significativos no desenvolvimento e uso da IA. Estes riscos se concentram em questões éticas, sociais e existenciais, exigindo monitoramento e regulamentação constantes (BOSTROM, 2016; HINTON, 2023a/2021). O debate filosófico amplia essa reflexão ao mostrar que a técnica, quando desvinculada de valores éticos, pode comprometer não apenas estruturas sociais, mas a própria essência da existência humana.

No campo social e econômico, a IA pode aprofundar desigualdades por meio da substituição massiva de empregos e da concentração de poder tecnológico em poucas corporações. Essa dinâmica pode intensificar o que Marx (1867/2004) chamou de alienação do trabalho, transformando a tecnologia em um novo mecanismo de exploração e exclusão. Além disso, sistemas de recomendação e algoritmos sociais frequentemente replicam preconceitos, reforçando padrões de discriminação. Kant (1785/2007), ao afirmar que o ser humano deve ser sempre tratado como fim em si mesmo, alerta contra o risco de os indivíduos serem reduzidos a meros dados estatísticos manipulados por máquinas.

No plano político, emergem riscos diretos à democracia e à segurança social. O uso da IA em manipulação política e produção de Fake News ameaça os processos democráticos e a autonomia da opinião pública (HINTON, 2023a). Ao problematizar a técnica, Heidegger (2007) já advertia que o maior perigo não está na máquina em si, mas no enquadramento instrumental que reduz a realidade a simples recurso. Esse risco se manifesta em práticas de vigilância em massa e no desenvolvimento de armas autônomas, que concentram poder global e reduzem a pluralidade da existência à lógica do controle.

Quanto ao risco existencial e perda de controle humano, pesquisadores como Bostrom (2016) e Hinton (2021/2023a) destacam a possibilidade de que sistemas superinteligentes escapem ao domínio humano, representando uma ameaça à sobrevivência da espécie. Sartre (1943/2014), ao enfatizar que o homem é "condenado a ser livre", oferece uma chave crítica: a delegação irrestrita de decisões à IA pode significar a renúncia da liberdade humana em favor de uma racionalidade maquínica[11]. Esse perigo é reforçado pela opacidade dos algoritmos e pelo desenvolvimento de linguagens internas não compreensíveis, criando uma espécie de *caixa-preta*[12] tecnológica que mina a autonomia da razão humana (THE GUARDIAN, 2025).

Hans Jonas (2006), com seu princípio responsabilidade, ressalta que a humanidade deve sempre considerar os efeitos de suas ações tecnológicas sobre as gerações futuras. Assim, a criação de sistemas autônomos, sem prudência e sem limites éticos, pode comprometer não apenas a justiça presente, mas também a própria continuidade da vida. Nesse sentido, Aristóteles (2001), ao defender a *phronesis* (prudência) como virtude prática, reforça que a aplicação da IA deve estar ancorada no equilíbrio entre conhecimento técnico e reflexão ética, evitando que a busca pelo poder técnico comprometa os fundamentos da existência.

Weber (1905/1930) denuncia a racionalização que subordina o indivíduo à lógica mecânica; Ghiraldelli (2021) mostra que o *General Intelect* se amplia em sinais, mas perde corpo e afetividade; e Berardi (2012), ao falar em *semiocapitalismo*, evidencia uma sociedade conectada, porém sem verdadeira união. Em conjunto, revelam a mesma crítica: a modernidade tecnológica cria eficiência e conexão, mas esvazia a experiência humana e o sentido compartilhado.

4.3 Síntese evolutiva e perspectivas futuras

A análise dos aspectos positivos e negativos da Inteligência Artificial evidencia que seu impacto sobre a evolução humana dependerá menos da tecnologia em si e mais da forma como será orientada, regulamentada e compreendida em seu contexto ético e social. A IA, longe de ser um agente

[11] Racionalidade maquínica — se refere a uma forma de subjetividade e pensamento moldada por máquinas, tecnologias e sistemas de controle que operam fora da consciência ou da razão autônoma do indivíduo. É caracterizada pela produção de respostas reflexas e automáticas, impulsionadas por imagens da mídia e redes sociais, que subordinam o sujeito à lógica da dominação capitalista e a um funcionamento em "servidão maquínica".

[12] Caixa preta — em ciência, computação e engenharia — significa um sistema cujo funcionamento interno é desconhecido ou irrelevante, sendo analisado apenas com base em suas entradas e saídas ou características de transferência, ou melhor, é aquele em que o usuário não consegue ver o funcionamento interno do algoritmo. O termo pode se aplicar a diversos sistemas, desde componentes eletrônicos até processos complexos. O oposto de uma caixa preta é uma caixa branca, onde o funcionamento interno é conhecido e pode ser analisado.

neutro, funciona como um espelho das escolhas humanas, podendo tanto ampliar capacidades cognitivas, sociais e existenciais quanto exacerbar desigualdades, alienação e riscos existenciais.

A literatura indica que um caminho evolutivo positivo está condicionado à articulação de três dimensões centrais:

- Regulamentação ética e direitos humanos: a criação de marcos normativos sólidos deve garantir que a IA seja utilizada de forma a respeitar a dignidade e a autonomia de todos os indivíduos, em consonância com princípios kantianos (KANT, 1785/2007) de tratamento do ser humano como fim em si mesmo e com o *Princípio Responsabilidade* de Jonas (2006).
- Desenvolvimento tecnológico responsável: a inovação deve priorizar segurança, transparência e mitigação de riscos, se alinhando às recomendações de Bostrom (2016), Russell (2019) e Hinton (2023a). A prudência aristotélica (*phronesis*) oferece aqui um referencial filosófico para equilibrar potencial criativo e cuidado com consequências futuras (ARISTÓTELES, 2001).
- Educação e conscientização social: preparar cidadãos e profissionais para o convívio crítico com sistemas inteligentes é fundamental. A IA deve ser integrada a práticas de aprendizado e reflexão que promovam liberdade existencial (SARTRE, 1943/2014) e combate à alienação social e laboral (MARX, 1867/2004), garantindo que o progresso tecnológico seja instrumento de emancipação e não de dominação.

Essa síntese reforça que a tecnologia, quando orientada por valores filosóficos sólidos, pode se tornar uma ponte para uma evolução coletiva mais ética, cognitiva e existencialmente significativa.

5 CONSIDERAÇÕES FINAIS

A Inteligência Artificial representa um avanço tecnológico de caráter irreversível, que surge como uma oportunidade sem precedentes para a evolução humana, ao mesmo tempo em que impõe riscos éticos, sociais e existenciais significativos. Turing forneceu o arcabouço teórico que fundamenta a inteligência das máquinas, e Hinton (2021/2023b) materializou essas possibilidades, transformando conceitos em tecnologias capazes de interferir diretamente na vida cotidiana. A trajetória da IA evidencia que seu impacto dependerá menos da tecnologia em si do que da forma como será orientada, regulamentada e integrada a valores humanos, exigindo vigilância ética contínua.

Quando guiada por princípios éticos sólidos, a IA pode potencializar a criatividade, a empatia e o conhecimento humano, contribuindo para a construção de sociedades mais justas e sustentáveis. Aristóteles (384 a.C. e 322 a.C /2001), ao defender a prudência como virtude prática, lembra que o uso da técnica deve ser equilibrado pelo discernimento e pela reflexão sobre as consequências das ações. Kant (1785/2007) reforça que cada indivíduo deve ser tratado como fim em si mesmo, assegurando que os algoritmos não reduzam seres humanos a meros instrumentos de cálculo.

Jonas (2006) propõe que a responsabilidade com as futuras gerações seja um critério central de toda inovação, e Sartre (1943/2014) enfatiza que a liberdade humana não pode ser delegada sem consciência, mantendo a autonomia diante da expansão tecnológica. Marx (1867/2004) alerta para o risco de alienação e concentração de poder, lembrando que a tecnologia, se mal orientada, pode reproduzir desigualdades e comprometer a justiça social, enquanto Heidegger (2007) ressalta que a técnica, se percebida apenas como recurso instrumental, pode obscurecer formas mais amplas de existência.

A questão norteadora desse ensaio foi respondida, conscientizado que o equilíbrio entre inovação tecnológica e responsabilidade ética se faz necessário para que a IA funcione como ponte entre o que somos e o que podemos nos tornar. A análise realizada demonstra que, se orientada por regulamentações robustas, valores humanísticos e compromisso coletivo, a IA pode corrigir falhas comportamentais e existenciais, preservar o conhecimento, apoiar a saúde, promover sustentabilidade e fomentar o desenvolvimento cognitivo e social da humanidade. Entretanto, a ausência de reflexão crítica, de governança responsável e de integração filosófica pode transformar a mesma tecnologia em fator de desigualdade, manipulação ou perda de controle sobre sistemas complexos, tornando urgente a vigilância ética e a educação social.

Em síntese, a Inteligência Artificial constitui um marco que redefine as possibilidades humanas. Seu avanço inexorável exige de nós responsabilidade, prudência e coragem para agir com consciência. Somente através da articulação entre valores éticos, reflexão filosófica e inovação tecnológica é possível transformar a IA em instrumento de evolução consciente, capaz de ampliar o potencial humano e assegurar que a tecnologia permaneça a serviço da vida, da liberdade e da continuidade positiva da humanidade.

Este ensaio não esgota o tema e apresenta limitações relacionadas à ausência de dados empíricos e estudos de caso recentes, além de não explorar questões jurídicas e econômicas em profundidade. Também não foram abordadas implicações psicológicas individuais do convívio

humano com inteligências artificiais, deixando espaço para pesquisas futuras que ampliem essa discussão.

Diante dos resultados deste ensaio, algumas questões relevantes para pesquisas futuras para evidenciar a inteligência artificial como uma ponte (elo) entre o que somos e o que podemos nos tornar — desde que tenhamos coragem de agir com consciência: Como desenvolver mecanismos globais de regulamentação da IA que sejam eficazes diante de interesses políticos e econômicos divergentes? De que forma a IA pode promover transformação social positiva sem acentuar desigualdades ou comprometer direitos humanos fundamentais?

REFERÊNCIAS

ARISTÓTELES. **Ética a Nicômaco**. Tradução de Antônio Pinto de Carvalho. São Paulo: Martin Claret, (384 a.C. e 322 a.C.) 2001.

BBC NEWS. **O 'padrinho' da inteligência artificial que se demitiu do Google e adverte sobre perigos da tecnologia.** [*online*]. Artigo Public. São Paulo: BBC *News* Brasil, 02 maio 2023, [n.p.]. Disponível em: <https://www.bbc.com/portuguese/articles/cgr1qr06myzo>. Acesso em: 30 ago. 2025.

BERARDI, Franco "Bifo". ***Alma y el trabajo, El***. Idioma espanhol. Tradução Fabrizio Cossalter e Pagliari. Barcelona: Edit. Educal, 2012. ISBN: 9786075161136,

BOSTROM, Nick. ***Superintelligence****: Paths, Dangers, Strategies*. Publicação em inglês. 1.a ed. *Oxford: Oxford University Press,* 2016, 432 p. ISBN: 9780198739838.

BUSINESS INSIDER. ***Godfather of AI' says he's 'glad' to be 77 because the tech probably won't take over the world in his lifetime***. [*online*]. [s.l.]: *Business Insider*, 2025, [n.p.]. Disponível em: <https://www.businessinsider.com/ai-godfather-geoffrey-hinton-superintelligence-risk-takeover-2025-4>. Acesso em: 03 ago. 2025.

FORBES. **Os 15 maiores riscos da inteligência artificial.** [*online*] [s.l.]: Forbes Tech, 05 jun. 2023, [n.p.]. Disponível em: <https://forbes.com.br/forbes-tech/2023/06/os-15-maiores-riscos-da-inteligencia-artificial/>. Acesso em: 25 jul. 2025.

GHIRALDELLI, Paulo. **O que é subjetividade maquínica?** [*online*] Campo Grande: Girarldelli.org, 07 out. 2021, [n.p.]. Disponível em: <https://ghiraldelli.org/2021/10/07/o-que-e-subjetividade-maquinica/>. Acesso em 18 jun. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; & COURVILLE, Aaron. ***Deep Learning***. Publicação em Inglés. *Cambridge*: *MIT Press*, 2016a, 775 p. ISBN: 9780262035613.

GOODFELLOW, Iva; BENGIO, Yoshua; & COURVILLE, Aaron. **Aprendizado profundo**. [*online*]. [s.l.]: *Deep Learning Book Org,* 2016b, [n.p.]. Disponível em: <https://www.deeplearningbook.org/>. Acesso em: 15 jun. 2025.

HEIDEGGER, Martin. **A questão da técnica**. Tradução de Emmanuel Carneiro Leão, Gilvan Fogel e Márcia Sá Cavalcante Schuback. In: HEIDEGGER, Martin. Ensaios e Conferências. Petrópolis: Vozes, 2007. p. 11-38.

HINTON, Geoffrey. **Hinton Geoffrey adverte sobre os perigos da Inteligência Artificial.** [*online*]. [s.l.]: *Dezaine*, 2021, [n.p.]. Disponível em: <https://www.dezaine.co.mz/inicio/d6i3n9g75pwvpmo5sjr4lasontk99u>. Acesso em: 01 ago. 2025.

HINTON, Geoffrey. **Entrevista à *BBC News* sobre riscos da IA**. [*online*]. Londres: BBC, 2023a, [n.p.].

HINTON, Geoffrey. **O 'padrinho' da IA que se demitiu do Google e alerta sobre perigos da tecnologia**. [*online*]. Londres: BBC, 2023b, [n.p.]. Disponível em: <https://www.bbc.com/portuguese/media-65722593>. Acesso em: 25 ago.2025.

PEDRÓ, Francesc; SUBOSA, Miguel; RIVAS, Axel; VALVERDE, Paula. ***Artificial Intelligence in Education***: *Challenges and Opportunities for Sustainable Development.* [*online*]. Idioma Inglês. ED-2019/WS/8. Paris: UNESCO, 2019, 31 p. Disponível em: <https://www.academia.edu/68859749/Artificial_intelligence_in_education_challenges_and_opportunities_for_sustainable_development>. Acesso em: 29 ago. 2025.

JONAS, Hans. **O princípio responsabilidade**: ensaio de uma ética para a civilização tecnológica. Tradução: Marijane Lisboa e Luiz Barros Montez. Rio de Janeiro: PUC-Rio; Contraponto, 2006, 356 p. ISBN: 978-85-85910-84-6.

KANT, Immanuel. "***Fundamentação da metafísica dos costumes***". Trad. Paulo Quintela. Lisboa: Edições 70, (1785) 2007, pp. 47-87.

MARX, Karl. **Manuscritos econômico-filosóficos**. Tradução de Jesus Ranieri. São Paulo: Boitempo Editorial, (1867) 2004, 176 p. ISBN: 978-8575590027.

PETRY, F. B. O conceito de razão nos escritos de Max Horkheimer. In: **Cadernos De Filosofia Alemã**: Crítica e Modernidade, 22 ed., São Paulo: Depto. Filosofia FFLCH-USP, 2013 pp. 31-48. Disponível em: <https://doi.org/10.11606/issn.2318-9800.v0i22p31-48>. Acesso em: 23 ago. 2025.

RUSSELL, Stuart. ***Human Compatible:*** *Artificial Intelligence and the Problem of Control*. Idioma inglês. *New York: Viking,* 2019, 352 p. ISBN: 9780525558613.

RUSSELL, Stuart; & NORVIG, Peter. ***Artificial Intelligence****: A Modern Approach*. Publicação em Inglês. *Pearson Series in Artificial Intelligence.* 4. ed. *Upper Saddle River*: Pearson, 2020, 1136 p. ISBN: 9780134610993.

SARTRE, Jean-Paul. **O existencialismo é um humanismo**. Tradução de João Batista Kreuch, 4. ed., Petrópolis: Vozes, (1943) 2014, 60 p.

SIEMENS, George. ***Connectivism: A learning theory for the digital age***. *International Journal of Instructional Technology and Distance Learning*. Idioma inglês. v. 2, n. 1, 2008, 59 p. Disponível em:<https://www.academia.edu/2857237/Connectivism_a_learning_theory_for_the_digital_age>. Acesso em: 25 ago. 2025.

TEGMARK, Max. ***Life 3.0****: Being Human in the Age of Artificial Intelligence.* [*online*] Pdf. *New York: Alfred Alfred A. Knopf*, 2017. 440 p. ISBN: 9781101946602. Disponível em: <https://www.cag.edu.tr/uploads/site/lecturer-files/max-tegmark-life-30-being-human-in-the-age-of-artificial-intelligence-alfred-a-knopf-2017-aTvn.pdf>. Acesso em: 25 jul. 2025.

THE GUARDIAN***. Godfather of AI' shortens odds of the technology wiping out humanity over next 30 years.*** [*online*]. Publicação em inglês. Australia: *Suport the Guardiam*, 2025, [n.p.]. Disponível em: <https://www.theguardian.com/technology/2024/dec/27/godfather-of-ai-raises-odds-of-the-technology-wiping-out-humanity-over-next-30-years >. Acesso em: 03 ago. 2025.

TURING ARCHIVE. ***Turing Quotes***: Alan Turing (1912-1954). [*online*]. Publicação em inglês. *Cambridge, UK: The Turing Digital Archive,* 2025, [n.p.]. Disponível em: <https://turingarchive.kings.cam.ac.uk/turing-quotes>. Acesso em: 21 ago. 2025.

TURING, Alan Mathison. ***Computing Machinery and Intelligence***. Publicação em inglês. *Mind,* Vol. LIX, *Issue* 236, *October* 1950, p. 433–460. Disponível em: <https://doi.org/10.1093/mind/LIX.236.433>. Acesso em: 15 jul. 2025.

WEBER, Max. ***The protestant ethic and the spirit of capitalism***. [*online*] Idioma inglês. *Trasnslated By* Talcott Parson. *Nova York: Charles Scribner's Sons*, (1905). 1930. Disponível em: <https://ghiraldelli.org/2021/10/07/o-que-e-subjetividade-maquinica/>. Acesso em: 25 ago.2025.

DATA DA ENTREGA DO TCC À BANCA PARA CONFERÊNCIA: **08/09/2025**

DIA DA APROVAÇÃO DO TCC PELA BANCA: **11/09/2025**

DATA EMISSÃO DA CARTA/DECLARAÇÃO DE APROVAÇÃO DO TCC: **04/10/2025**

RESPONSÁVEL PELA ATESTAÇÃO: **SEBASTIÃO FLÁVIO MOTIM DA SILVA - Diretor geral FAAL**