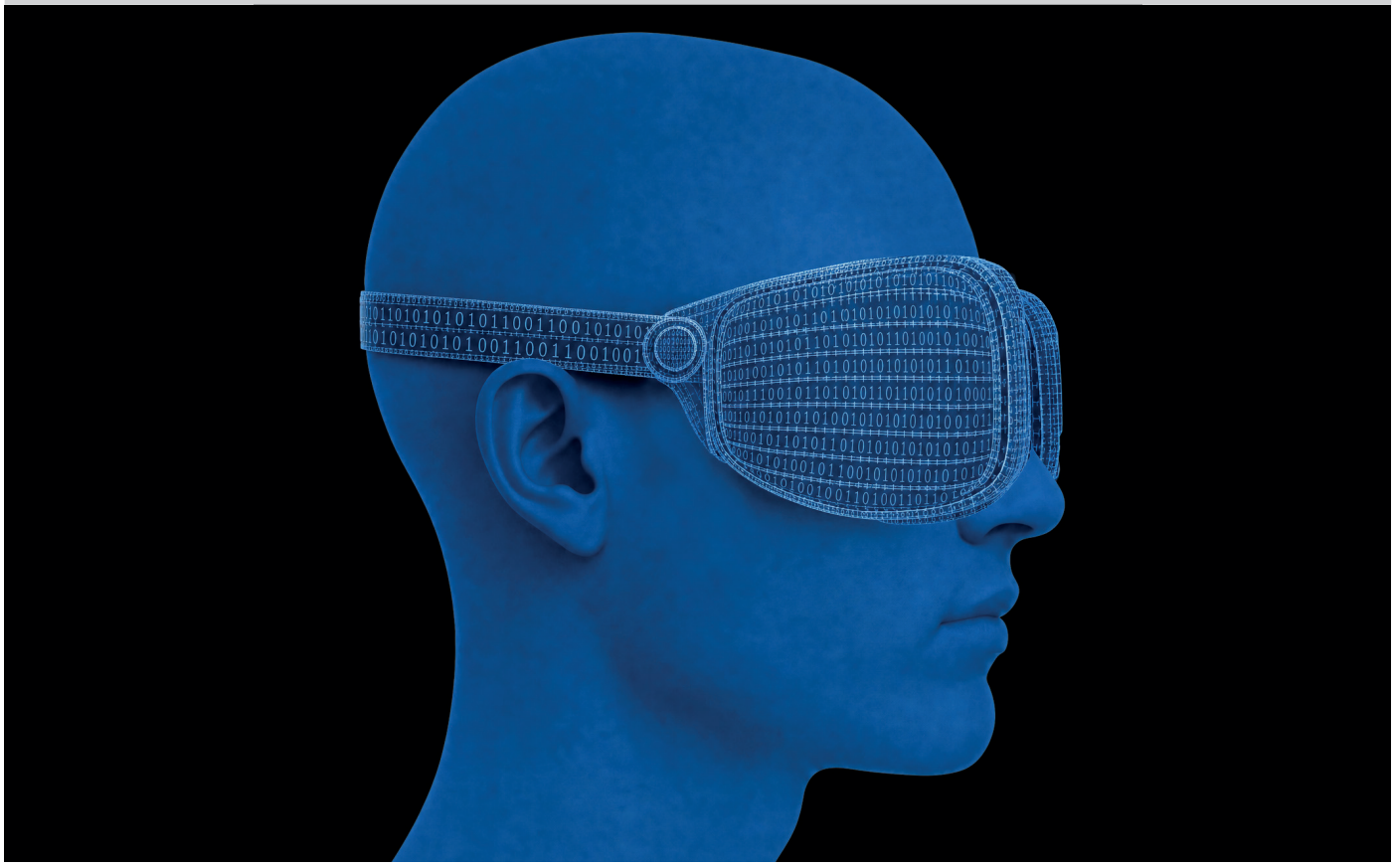
EFEITOS COLATERAIS DA INTELIGÊNCIA ARTIFICIAL: IMPLICAÇÕES PARA INDIVÍDUOS, ORGANIZAÇÕES E SOCIEDADE

SIDE EFFECTS OF ARTIFICIAL INTELLIGENCE: IMPLICATIONS FOR INDIVIDUALS, ORGANIZATIONS, AND SOCIETY



O uso da IA traz ganhos de produtividade e inovação, mas também implica riscos éticos, cognitivos e estruturais. Este artigo organiza os principais efeitos colaterais identificados na literatura recente e sugere caminhos para mitigar seus impactos em diferentes níveis

The use of AI brings gains in productivity and innovation, but also implies ethical, cognitive, and structural risks. This article organizes the main side effects identified in recent literature and suggests ways to mitigate their impacts at different levels

Suely Fischer Omura | sfomura17@gmail.com | ORCID: 0000-0003-1887-7700
Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, SP, Brasil

Vanessa Martins dos Santos | vanessadossantos2611@gmail.com | ORCID: 0000-0003-4330-578X
Escola Superior de Propaganda e Marketing, São Paulo, SP, Brasil
FIA Fundação Instituto de Administração, São Paulo, SP, Brasil

Resumo

Objetivo: propor *framework* dos efeitos colaterais decorrentes do uso da inteligência artificial (IA) e estratégias de mitigação de risco que permitam a diferentes atores reduzir os impactos éticos, cognitivos e estruturais associados à tecnologia.

Estado da arte: discussão sobre IA evolui de um enfoque técnico, centrado em desempenho e transparência, para uma perspectiva humanista, preocupada com o bem-estar, a integridade cognitiva e a responsabilidade social no uso da tecnologia.

Escopo: foi realizada revisão sistemática da literatura com artigos publicados entre 2023 e 2024, em periódicos de alto impacto, identificando efeitos adversos da IA e diretrizes de mitigação.

Originalidade: o estudo expõe os efeitos colaterais decorrentes do uso de ferramentas de IA, relaciona o tipo de risco ao nível de impacto e propõe ações de mitigação desses riscos para garantir o uso ético, responsável e eficaz da IA.

Impacto: o *framework* proposto orienta ações práticas e políticas para o uso ético, responsável e eficaz da IA.

Palavras-chave: efeitos colaterais, uso de IA, riscos desafios relacionados à IA, mitigação de riscos.

ODS: 3 – Saúde e bem-estar, 4 – Educação de qualidade, 8 – Trabalho decente e crescimento econômico, 9 – Indústria, inovação e infraestrutura, 16 – Paz, justiça e instituições eficazes.

ABSTRACT

Objective: *To propose a framework addressing the side effects arising from the use of Artificial Intelligence (AI) and risk mitigation strategies that enable different actors to reduce the ethical, cognitive, and structural impacts associated with this technology.*

State of the Art: *The discussion on AI is evolving from a technical focus, centered on performance and transparency, to a humanistic perspective concerned with well-being, cognitive integrity, and social responsibility in technology use.*

Scope: *A systematic literature review was conducted of articles published between 2023 and 2024 in high-impact journals, identifying adverse effects of AI and mitigation guidelines.*

Originality: *The study highlights the side effects of AI tools use, relates the type of risk to its level of impact, and proposes mitigation actions to ensure the ethical, responsible, and effective use of AI.*

Impact: *The proposed framework guides practical and policy actions for the ethical, responsible, and effective use of AI.*

Keywords: *side effects, AI use, AI-related challenges risks, risk mitigation.*

SDG: *3 – Good Health and Well-Being; 4 – Quality Education; 8 – Decent Work and Economic Growth; 9 – Industry, Innovation and Infrastructure; 16 – Peace, Justice and Strong Institutions.*



Dois arquétipos da psicologia junguiana – anjo bom (princípio orientador) e anjo mau (impulso destruidor) – dialogam entre si na mente de um executivo:

Anjo bom: – Você precisa apresentar o projeto de reestruturação do portfólio na próxima reunião. Vamos trabalhar, o tempo está acabando, e você não domina esse tema.

Anjo mau: – Não se preocupe. Você pode usar a IA como atalho, mas não revele a fonte, assim todos se surpreenderão com sua performance singular.

Anjo bom: – Cuidado, você não tem como garantir que a IA vai gerar insights precisos e confiáveis. Pior, se você não creditar o conteúdo à IA, você corre o risco de ser questionado por comportamento antiético.

Reflexão do executivo: – E agora, que faço? Como garantir um uso seguro?

Apesar de o diálogo constituir uma caricatura da realidade, ele espelha de modo eficaz a divisão de águas existente na sociedade no que se refere ao uso da IA. De um lado, impera uma overdose dos aspectos positivos da adoção de ferramentas baseadas em IA e, do outro, cresce uma vertente que enaltece os aspectos controversos da tecnologia e potenciais ameaças decorrentes do seu uso. Essa visão dicotômica demonstra a urgência de se estabelecer um equilíbrio no cenário, ajustando as vantagens da IA e as ameaças implícitas no seu uso a um nível mais realista e separando estereótipos pessoais de riscos autênticos e justificados. Dessa forma, é possível alcançar um uso disruptivo, mas simultaneamente ético, responsável e eficaz[1].

Por serem capazes de executar tarefas que previamente requeriam inteligência e funções cognitivas humanas, *chatbots* de IA generativa prometem revolucionar o nosso trabalho e nossa vida, transformando tarefas complexas em atividades elementares. No entanto, apesar das inúmeras vantagens oferecidas, como maior acessibilidade, velocidade e eficiência no processo decisório e na transformação de processos e operações de negócios, despontam também apreensões com relação a potenciais efeitos colaterais atrelados ao seu uso, como desinformação[3], conteúdo fabricado ou tendencioso[4], disrupção de relações humanas[5], dependência cognitiva[6], redução da autonomia humana[7], riscos de segurança e privacidade de dados[8], e até mesmo relacionadas à saúde mental e bem-estar dos usuários[9].

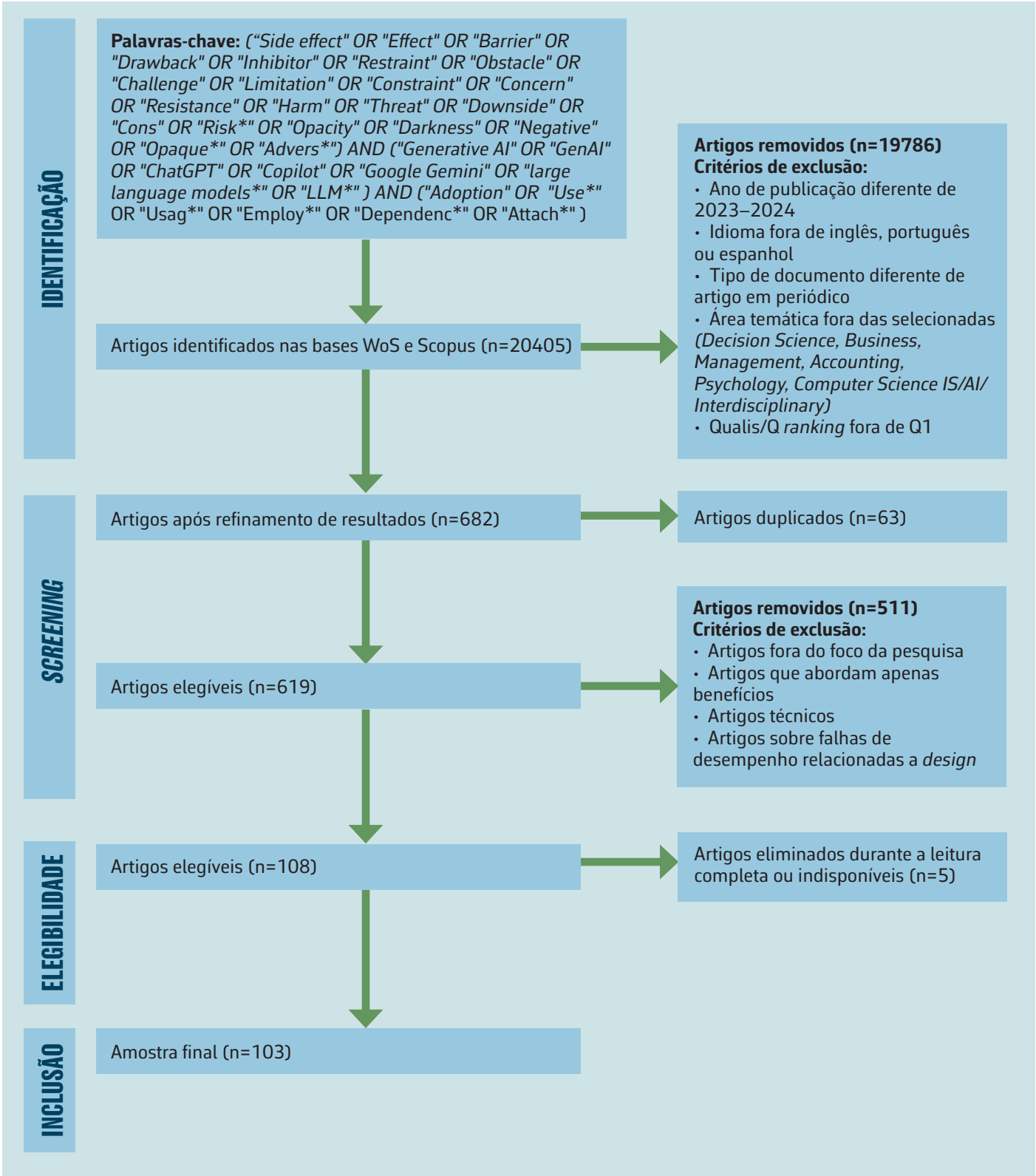
Esse foco crescente dos estudos acadêmicos nos riscos inerentes à IA generativa e suas implicações nos motivou a buscar na literatura recente as desvantagens do uso dos aplicativos baseados em IA e os efeitos colaterais mais acentuados, além de diretrizes de mitigação dessas consequências desfavoráveis para promover um uso mais consciente, ético e responsável. A partir da análise, o artigo identifica e organiza os principais efeitos colaterais da IA em diferentes níveis — individual, organizacional e societal — e propõe recomendações de mitigação direcionadas aos diferentes atores envolvidos em seu desenvolvimento, regulação e uso.

METODOLOGIA

Foi realizada revisão sistemática da literatura, com pesquisa de palavras-chave relacionadas a efeitos colaterais, barreiras, riscos e limitações no uso de IA generativa, com foco em tecnologias baseadas em *large language models* (LLMs), como ChatGPT, Copilot e Gemini. As combinações de termos incluíram expressões referentes aos efeitos colaterais da tecnologia associadas a descritores de adoção, uso e dependência dessas ferramentas em contextos organizacionais e individuais, buscando capturar estudos que abordassem as consequências, limitações e resistências associadas ao uso de IA generativa.

Figura 1.

Prisma



A busca foi realizada nas bases Web of Science e Scopus, considerando artigos publicados entre 2023 e 2024 e redigidos em inglês, português ou espanhol. Esse recorte temporal foi escolhido porque o período marca a consolidação e difusão global das ferramentas de IA generativa, após o lançamento público do ChatGPT e de outros sistemas similares. A partir de 2023, observou-se um aumento significativo na produção científica voltada ao tema, justificando o foco analítico nesse intervalo. Após filtragem (Figura 1), a amostra foi composta de 103 artigos revisados por pares.

Cada artigo foi codificado segundo um protocolo analítico estruturado, contemplando as seguintes categorias: título, contexto de aplicação, tecnologia de IA utilizada, descrição técnica, tipo de uso, efeitos colaterais identificados, alvo e tipo de efeito colateral, informações complementares (como conclusões e observações teóricas), método de pesquisa, questão de pesquisa, implicações práticas, direções para pesquisas futuras e resultados, além de variáveis descritivas sobre amostra populacional, objetivos, achados e resumo do artigo.

A etapa inicial da codificação contou com o apoio da ferramenta de IA SciSpace, utilizada para gerar automaticamente os descritores padronizados nas colunas correspondentes da matriz analítica. O SciSpace atuou apenas na organização preliminar dos descritores; todas as decisões analíticas e as validações foram realizadas manualmente pelas autoras, assegurando que o processo permanecesse fundamentado em julgamento humano especializado. Em seguida, foi conduzida uma revisão manual detalhada, com leitura integral dos textos para conferência das informações e complementação das colunas que não são cobertas pela ferramenta, por exemplo, contexto, uso da IA e tipo de efeito colateral. A categorização por contextos é apresentada na Tabela 1.

Tabela 1

Classificação da literatura por contexto

N.	Domínio	Número de artigos
1	Ambientes Educacionais	39
2	Setor Empresarial	20
3	Negócios e Plataformas *On-line*	15
4	Múltiplos Campos	13
5	Serviços Profissionais	10
6	Indústrias Criativas/Entretenimento/*Design* de Moda e Vestuário	4
7	Hospedagem e Turismo	2

ESTADO DA ARTE

O avanço da IA redefiniu as fronteiras entre humano e tecnológico. Ao mesmo tempo que amplia a eficiência e a capacidade de análise, a IA introduz efeitos colaterais que se espalham por dimensões éticas e legais, cognitivas e emocionais ou estruturais, afetando indivíduos, organizações e a sociedade de maneira interconectada[10]. Esses efeitos não surgem isoladamente: eles formam uma rede complexa em que o uso não crítico da IA pode corroer a confiança, fragilizar instituições, gerar dependência emocional e aprofundar desigualdades[11]. Esses desafios, se bem geridos, proporcionam benefícios a diversos *stakeholders*. No entanto, se malconduzidos, produzem potenciais efeitos colaterais que também impactam múltiplos *stakeholders* e, portanto, requerem planos de ação para mitigação de riscos.

Em contextos educacionais e científicos, a automação da escrita e da produção intelectual desafia princípios fundamentais de integridade e autoria[12]. A facilidade de gerar textos, dados e imagens automatizados estimula práticas de plágio, falsificação e fraude acadêmica, diluindo o valor da originalidade e da reflexão crítica[13]. Quando o conteúdo perde autenticidade, o conhecimento deixa de ser construído e passa apenas a ser reproduzido, com menor rigor, menor contribuição do autor humano e menor capacidade de inovação[14]. Essa erosão intelectual conecta-se ao declínio de desempenho humano e ao empobrecimento das competências cognitivas, fenômenos que também se estendem ao ambiente organizacional[15].

O uso intensivo da IA altera o modo como as pessoas aprendem, trabalham e se relacionam com o próprio saber. A terceirização do pensamento crítico e da criatividade leva a uma crescente obsolescência de competências, à redução do engajamento e à padronização excessiva de respostas[16]. Ao depender de soluções automáticas, o profissional corre o risco de perder tanto o domínio técnico quanto o entusiasmo pela resolução autônoma de problemas. A automatização do julgamento, no lugar da automatização das tarefas, ou melhor, do trabalho da IA como assistente do ser humano, gera apatia cognitiva e insegurança sobre o próprio valor humano no processo produtivo[17].

Esses efeitos mentais e comportamentais são acompanhados por reações emocionais intensas. O medo de substituição por sistemas inteligentes e o fluxo incessante de estímulos digitais provocam ansiedade, estresse ocupacional e sobrecarga tecnológica, reduzindo o bem-estar e ampliando quadros de exaustão mental[18]. Paralelamente, a dependência emocional e cognitiva dos sistemas de IA faz com que o sujeito deixe de confiar em si mesmo e passe a delegar à máquina sua capacidade de decidir, planejar e até interpretar o mundo[19].

Essa relação de confiança atribuída à tecnologia (IA), e não à inteligência humana, também se manifesta no plano institucional. Organizações que adotam a IA de maneira indiscriminada correm o risco de erodir seus fluxos de conhecimento, substituindo a experiência humana por bases de dados e algoritmos opacos[20]. O resultado é a desintegração organizacional, com a perda de rigor, enfraquecimento das práticas profissionais e deslocamento da autoridade decisória para sistemas automatizados. O que se apresenta como eficiência técnica pode, na prática, gerar fragilidade cognitiva e dependência estrutural[21].

A falta de transparência desses sistemas aprofunda o problema. Modelos de "caixa-preta", desprovidos de empatia ou sensibilidade cultural, reproduzem vieses de gênero, raça e classe, amplificando desigualdades históricas e distorções decisórias. O viés algorítmico não é apenas técnico; é o resultado das assimetrias sociais refletidas nos dados que são utilizados para treinamento da IA[22]. Assim, dilemas éticos e de justiça emergem, exigindo governança, diversidade nas equipes de desenvolvimento e mecanismos de proteção[23].

Ao mesmo tempo, cresce a preocupação com manipulação e má conduta no uso da IA. *Deepfakes*, desinformação e conteúdos fabricados impactam a confiança[24]. As mesmas tecnologias que permitem ganhos de produtividade podem ser usadas para enganar, distorcer ou influenciar comportamentos de maneira antiética[25]. Essa manipulação soma-se a novas formas de assédio e comportamentos nocivos nos ambientes digitais, que incluem a amplificação de discursos de ódio, exploração sexual e processos de radicalização[26]. A IA, quando não regulada, não apenas reproduz violência simbólica; ela a potencializa, convertendo algoritmos em vetores de agressividade social[27].

A confiança, portanto, emerge como uma das variáveis centrais na relação humano-máquina. A perda de credibilidade e autenticidade em informações, marcas e interações digitais reduz o engajamento e mina relações institucionais[28]. Esse fenômeno está diretamente ligado à ausência de *accountability* e transparência nos processos automatizados. Sem mecanismos claros de responsabilização, a fronteira entre erro técnico e decisão ética torna-se difusa[29].

Outro campo de risco é o da privacidade e segurança de dados. A coleta massiva de informações, os vazamentos e a vigilância algorítmica configuram um novo tipo de vulnerabilidade, que combina ameaças técnicas e éticas[30]. A exposição de dados pessoais e sensíveis fragiliza a autonomia dos indivíduos e desafia a governança organizacional. De modo mais amplo, tais práticas contribuem para a concentração de poder em grandes corporações tecnológicas e para o aumento das assimetrias informacionais[31].

No nível societal, os impactos ambientais e sistêmicos completam o quadro de efeitos colaterais. A IA demanda alta capacidade computacional, consumo energético e infraestrutura física intensiva, o que pressiona recursos naturais e amplia a pegada de carbono digital[32]. Além disso, a centralização do poder técnico e a automatização de decisões econômicas e políticas criam novas dependências sistêmicas[33].

Esses achados foram condensados na Tabela 2, que organiza os efeitos identificados e serve de base direta para a estruturação da matriz proposta (Figura 2).

Tabela 2.

Principais efeitos colaterais

Tópico	Principais Efeitos
Impactos ambientais e sistêmicos (44)	Transformações no trabalho e nas estruturas econômicas,; redefinição da agência humana e das práticas institucionais, desafios de governança e centralização do poder técnico, impactos ambientais e de sustentabilidade.
Limitações cognitivas e de dificuldades de compreensão (39)	Declínio do pensamento crítico e do discernimento, limitação cognitiva geral, aprendizado superficial e desengajamento cognitivo, empobrecimento criativo e expressivo, redução da inteligência e da reflexividade emocional, risco existencial e desumanização.
Declínio de desempenho (38)	Desmotivação e redução do engajamento, obsolescência e degradação de competências, queda da produtividade e do desempenho técnico.
Vieses e justiça (36)	Vieses técnicos em dados e algoritmos, desinformação e respostas enviesadas, vieses cognitivos humanos, discriminação e injustiça sistêmica.
Ansiedade, estresse e bem-estar emocional (35)	Estresse ocupacional e perda de bem-estar, comprometimento da saúde física e mental, sofrimento emocional e psicológico, ansiedade e tecnossobrecarga.
Manipulação e má conduta (35)	Distorção e fabricação de informações, manipulação comportamental e enganosa, uso indevido da IA e outras condutas antiéticas.
Questões éticas, legais e sociais (33)	Desigualdade de acesso e exclusão digital, dilemas éticos, culturais e morais, desafios de governança e justiça algorítmica.
Questões de integridade acadêmica (de conteúdo) (31)	Fraudes e trapaças acadêmicas, perda de autenticidade e originalidade, falsa autoria e falta de atribuição, violações legais ou de propriedade intelectual, comprometimento da integridade de avaliações, conteúdo falsificado ou não confiável.
Privacidade, segurança e riscos de dados (25)	Violações de privacidade e uso indevido de dados, vulnerabilidades de sistemas e infraestruturas, riscos de vigilância e monitoramento.
Problemas de confiança e responsabilidade (22)	Perda de confiança e credibilidade, dificuldades de responsabilização e falta de transparência.
Limitações de *design* (falta de empatia, compreensão cultural e caixa-preta) (17)	Opacidade e falta de explicabilidade, falta de empatia e de entendimento contextual, geração de conteúdo falso ou inconsistente (alucinações), dependência estrutural de dados e modelos.
Dependência e uso excessivo (15)	Dependência tecnológica e perda de autonomia cognitiva, dependência comportamental, dependência emocional e psicológica.
Disrupções nas relações humanas (14)	Isolamento, alienação e distanciamento social, redefinição de vínculos e interações humanas, desumanização e erosão do valor humano, desigualdade e divisão social.
Erosão do conhecimento e das organizações (12)	Desintegração do conhecimento organizacional, perda de rigor e de controle de qualidade, reconfiguração hierárquica e perda de autoridade institucional.
Assédio e comportamentos nocivos (6)	Comportamentos abusivos e violentos, radicalização e estigmatização social.

Nota. O número entre parênteses indica a quantidade de ocorrências de efeitos colaterais identificados nos artigos analisados e posteriormente agrupados em cada tópico temático.

Esses riscos interligados — éticos/legais, cognitivos/emocionais e estruturais — configuram um ecossistema de desafios que se retroalimentam. A falta de empatia dos sistemas, o isolamento nas relações humanas, a desigualdade digital, o declínio das competências e a erosão institucional formam um ciclo que ameaça tanto a autonomia individual quanto a sustentabilidade das organizações e da sociedade. Em última instância, os efeitos colaterais da IA não são apenas consequências técnicas, mas sintomas de uma transição civilizatória em que o humano precisa redefinir seu papel diante da máquina[34].

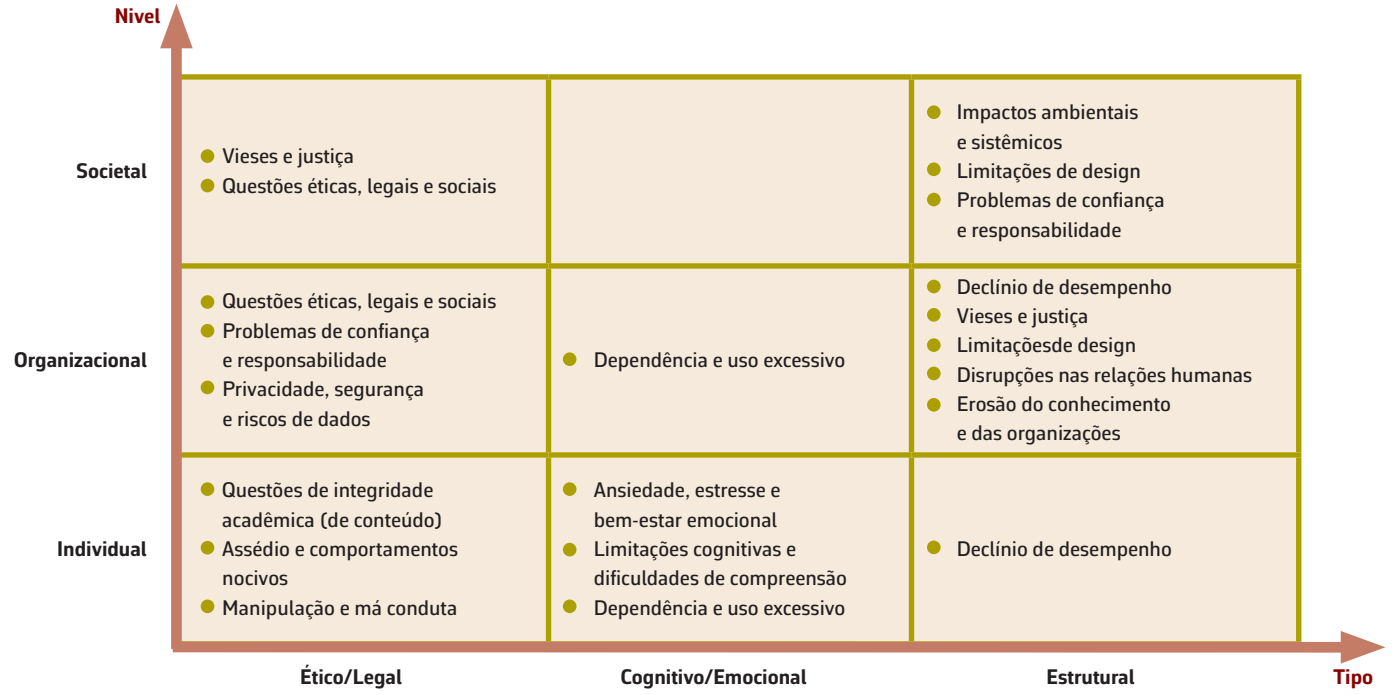
A revisão revelou três tipos recorrentes de risco (ético/legal, cognitivo/emocional e estrutural) distribuídos de modo consistente nos níveis individual, organizacional e societal. Esses padrões, observados de maneira transversal à literatura, sustentam a categorização adotada e orientam a formulação do *framework*.

Com base na síntese da Tabela 2, a matriz 3×3 organiza os efeitos segundo os tipos de risco e os níveis de impacto identificados na revisão, integrando os padrões recorrentes observados na literatura. Essa abordagem evidencia como preocupações aparentemente isoladas, como plágio acadêmico, ansiedade ou viés algorítmico, conectam-se a problemas mais amplos de confiança organizacional, desigualdade digital e impactos ambientais[35]. A matriz oferece uma visão integrada e prática para análise dos efeitos colaterais, permitindo que gestores, pesquisadores e formuladores de políticas compreendam melhor a extensão e a interconexão desses fenômenos.

Ao reconhecer que riscos éticos e cognitivos individuais podem evoluir para dilemas estruturais e societais, essa perspectiva propõe uma governança mais humanizada, transparente e sustentável da IA, reafirmando o princípio de que toda tecnologia deve permanecer a serviço das pessoas, e não o contrário.

Figura 2.

Matriz nível de impacto x tipo de risco



Nota. Uma categoria pode aparecer em mais de um campo da matriz.

DA TEORIA À AÇÃO: COMO MITIGAR OS IMPACTOS NEGATIVOS DA IA

A progressiva exposição pública dos efeitos adversos do uso de ferramentas baseadas em IA vem desencadeando um debate crescente sobre a necessidade de estabelecer estratégias de mitigação dos impactos negativos, preservando a saúde mental e o bem-estar dos usuários[36]. A Tabela 3 ilustra recomendações de ações corretivas que podem ser lideradas por usuários finais, organizações, instituições educacionais, órgãos reguladores e sociedade.

Entre as ações recomendadas mais proeminentes na literatura, temos as que visam minimizar os danos e a resistência à adoção de IA[37], promover a inclusão digital[38], garantir a relevância e precisão do conteúdo gerado[39], integrar e otimizar o uso da IA em ambientes educacionais[40], garantir o desenvolvimento e uso responsável da IA[41], assegurar a qualidade, privacidade e segurança dos dados[42], e preparar os usuários para os desafios da automação cognitiva por intermédio da literacia em IA[43]. Por meio desse conjunto de ações, com exemplos de práticas implementadas no mercado[44], diferentes *stakeholders* podem aferir se o desenvolvimento, integração e uso de ferramentas baseadas em IA estão em conformidade com os aspectos éticos, legais e sociais requeridos pela sociedade na atualidade e promover ajustes em caso de desalinhamento.

Tabela 3. (Continua)

Recomendações de ações corretivas

Nível de Impacto	Ético/Legal	Cognitivo/Emocional	Estrutural
Societal	• Vieses e justiça (36) • Questões éticas, legais e sociais (33)		• Impactos ambientais e sistêmicos (44) • Problemas de confiança e responsabilidade (22) • Limitações de design (falta de empatia, compreensão cultural e caixa-preta) (17)
Sociedade (comunidade organizada de seres humanos que compartilham normas, instituições e relações)	• Criar órgãos reguladores multisetoriais para desenvolver e integrar práticas éticas de IA (7) • Combater a desinformação e discriminação por meio de campanhas educativas em diversas plataformas, mídia impressa e campanhas em instituições educacionais públicas e privadas (4)	–	• Promover conferências e fóruns públicos com líderes da indústria, pesquisadores, órgãos reguladores, grupos de defesa de direitos e sociedade civil (11) • Promover workshops e cursos para preservar agência humana, aprofundar o entendimento ético da IA, promover o uso responsável e estimular a colaboração ativa entre humanos e IA na geração de informações (8) • Conscientizar e educar sobre os benefícios, riscos e limitações da IA por meio de campanhas públicas usando linguagem acessível, exemplos práticos, mídias sociais, vídeos e podcasts (4)
Órgãos Reguladores (formuladores de políticas, agências de supervisão e instituições reguladoras)	• Estabelecer normas, legislação, frameworks de governança, órgãos de auditoria e fiscalização, diretrizes de boas práticas para uso responsável da IA (20)	–	• Criar proteções legais, mecanismos de controle e estruturas de responsabilização com auditoria, documentação e credenciamento para sistemas de alto risco com diferentes camadas de stakeholders (6) • Atualizar as leis de direitos autorais para acomodar formas híbridas de criatividade, distinguir entre responsabilidade e autoria, exigir transparência sobre bases de dados e documentação da interação humano-IA (instruções dadas à IA, detalhes de intervenção humana, modificações feitas) (2)

Tabela 3. (Continua)

Recomendações de ações corretivas

Nível de Impacto	Ético/Legal	Cognitivo/Emocional	Estrutural
Individual	• Questões de integridade acadêmica (de conteúdo) (31) • Assédio e comportamentos nocivos (6) • Manipulação e má conduta (35)	• Ansiedade, estresse e bem-estar emocional (35) • Limitações cognitivas e dificuldade de compreensão (39) • Dependência e uso excessivo (15)	• Declínio de desempenho (38)
Usuários (funcionários, estudantes, professores, pesquisadores, pacientes, médicos, consumidores)	• Manter-se atualizado dos riscos e limitações da IA antes de delegar decisões, por meio de consultas às literaturas acadêmica e prática (3) • Assegurar integridade e transparência, atribuindo créditos à IA na geração de conteúdo ou no suporte à execução de tarefas (3) • Evitar usos impróprios ou antiéticos de IA e adotar conduta responsável e de engajamento em políticas públicas de boas práticas de adoção de IA (4)	• Desenvolver fluência em IA e diversificar habilidades por meio de entendimento de conceitos, do uso eficaz de prompts e workflows, da adoção de uma mentalidade de aprendizado contínuo, e da compreensão de que a colaboração humano-IA é mais benéfica do que a mútua exclusão (3) • Evitar sobrecarga cognitiva do uso excessivo de IA, alternando tarefas com e sem IA, desenvolvendo literacia digital emocional (compreender limites e vieses e racional usado pela IA), criando rotinas saudáveis e preservando interações sociais (8)	• Validar os *insights* de IA para garantir relevância e precisão do conteúdo gerado (5) por meio de um check-list que abrange confirmar a relevância das fontes, avaliar fluxo lógico, verificar precisão numérica, validar reconhecimento de padrões e verificar realidade específica do setor, entre outros aspectos (7)
Organizacional	• Questões éticas, legais e sociais (33) • Problemas de confiança e responsabilidade (22) • Privacidade, segurança e riscos de dados (25)	• Dependência e uso excessivo (15)	• Declínio de desempenho (38) • Vieses e justiça (36) • Limitações de design (falta de empatia, compreensão cultural e caixa-preta) (17) • Disrupções nas relações humanas (14) Erosão do conhecimento e das organizações (12)
Instituições Educacionais e Educadores (escolas, universidades, faculdades, centros de treinamento, mentores, professores, acadêmicos, orientadores)	• Estabelecer políticas, diretrizes e manuais para usos permitidos e proibidos de IA, integridade acadêmica, governança, treinamento, privacidade e uso seguro, com revisão contínua (30)	• Incorporar alfabetização em IA nos currículos, integrando competências técnicas, aplicações de negócios, ética, experiência prática e capacitação docente (12)	• Adaptar continuamente os métodos de ensino, as técnicas de aprendizado e a avaliação dos avanços das tecnologias de IA por meio do uso de plataformas de aprendizagem adaptativa (personalizar currículos, ritmo e apresentação de conteúdo), treinamento contínuo para educadores (chatbots e sistemas de tutoria inteligente para prover ajuda sob demanda e ambientes virtuais para aprimorar as disciplinas práticas) e uso de ferramentas de IA para avaliações adaptativas que ajustam a dificuldade das questões, analisam padrões de erros e preveem o desempenho para *feedback* personalizado (11)

Tabela 3. (Conclusão)

Recomendações de ações corretivas

Nível de Impacto	Ético/Legal	Cognitivo/Emocional	Estrutural
Organizacional	• Questões éticas, legais e sociais (33) • Problemas de confiança e responsabilidade (22) • Privacidade, segurança e riscos de dados (25)	• Dependência e uso excessivo (15)	• Declínio de desempenho (38) • Vieses e justiça (36) • Limitações de design (falta de empatia, compreensão cultural e caixa-preta) (17) • Disrupções nas relações humanas (14) Erosão do conhecimento e das organizações (12)
Instituições Educacionais e Educadores (escolas, universidades, faculdades, centros de treinamento, mentores, professores, acadêmicos, orientadores)	• Garantir acesso equitativo à IA, capacitação inclusiva, ambientes de aprendizagem eficazes (plataformas adaptativas, salas de aula virtuais e ferramentas de acessibilidade para diversos estilos de aprendizagem), treinamento em alfabetização em IA e experiências de aprendizagem personalizadas e métricas de impacto sobre aprendizagem e equidade (7)	• Desenvolver criatividade, escrita, resolução de problemas com estímulo à colaboração entre humanos e IA e avaliação crítica dos resultados - alguns recursos incluem sistemas de tutoria com tecnologia GenAI para orientação personalizada em tempo real, programa piloto para introduzir a IA em disciplinas, estabelecimento de métricas para habilidades (5) • Monitorar integridade acadêmica e dependência excessiva combinando ferramentas de detecção de IA, avaliações autênticas (projetos, simulações e tarefas colaborativas), políticas de transparência e treinamento de educadores (4)	• Integrar IA personalizada para acomodar diferentes conhecimentos e necessidades tecnológicas entre os alunos por meio de sistemas de aprendizagem adaptativa que personalizam o conteúdo educacional interativo e inclusivo (gamificação, e-books interativos, vídeos e plataformas compatíveis com dispositivos móveis para envolver diferentes estilos de aprendizagem e aumentar a acessibilidade), o ritmo e o *feedback* ao perfil único de cada aluno, aprimorando o engajamento, a acessibilidade e os resultados de aprendizagem (4)
Desenvolvedores de IA (desenvolvedores de software, cientistas/ engenheiros de dados, pesquisadores de IA, startups, empresas de tecnologia, comunidades de código aberto, institutos de pesquisa)	• Criar design de IA colaborativo e ético para assegurar inclusão digital, responsabilidade moral e conformidade com as leis e a sustentabilidade por meio de design centrado no humano, enfatizando princípios como justiça, responsabilidade e transparência, alinhamento com padrões como ISO/IEC 42001 e da incorporação de parâmetros de eficiência (IA orientada a objetivos, minimização de dados via aprendizado por transferência, priorização de energia renovável para infraestrutura e técnicas para reduzir o consumo de energia) (12)	• Incorporar métricas de bem-estar, entrevistas virtuais com perguntas comportamentais (treinamento e simulação dos modelos com input humano da informação comportamental a ser detectada pela IA), mecanismos de redução de vieses cognitivos e de uso compulsivo, e mecanismos para detectar violação de direitos autorais por meio de blockchain (5)	• Aprimorar a confiabilidade dos resultados por meio de prompts estruturados, mecanismos para salvar metadados, integração com outras tecnologias, apuração de identidade digital, validação de autenticidade de imagem (analisar inconsistências em nível de pixel, metadados, iluminação e plausibilidade da cena), identificação de regiões de alucinação e geração de outputs com base em evidências (9) • Desenvolver interfaces capazes de captar emoções, comportamento, percepções e necessidades por meio da integração de tecnologias multimodais de detecção de emoções e análise comportamental que permitem personalizar o conteúdo e oferecer respostas empáticas (6) • Reforçar segurança dos dados, transparência, justiça, interatividade e explicabilidade por meio de visibilidade das operações e IA explicável (ferramentas de visualização, demonstrações de raciocínio passo a passo e explicações em linguagem natural) e abordagens de segurança em múltiplas camadas (7)

CONSIDERAÇÕES FINAIS

A matriz de efeitos colaterais proposta sintetiza as principais implicações negativas do uso da IA identificadas na literatura recente e oferece um instrumento prático para análise e mitigação de riscos. A partir dela, indivíduos, organizações e instituições podem avaliar o quanto estão expostos a impactos éticos, cognitivos e estruturais e adotar estratégias que equilibrem inovação e responsabilidade. A matriz também atua como instrumento de diagnóstico, ao evidenciar onde os riscos se manifestam e orientar intervenções em contextos organizacionais, acadêmicos e educacionais, reforçando sua aplicabilidade para apoiar decisões e mitigar os efeitos adversos da IA generativa. Além disso, distingue-se de tipologias existentes ao integrar tipos de risco e níveis de impacto em um único instrumento, permitindo compreender a progressão dos efeitos entre dimensões individuais, organizacionais e societais.

Os efeitos mapeados variam em intensidade e duração, mas todos revelam uma transição em curso: a tecnologia deixa de ser apenas meio e passa a moldar a forma como pensamos, decidimos e nos relacionamos45. Nesse cenário, promover equilíbrio entre eficiência e humanidade torna-se essencial para que a IA amplie, e não substitua, a inteligência humana. Mais do que um alerta, o estudo propõe ações corretivas e *frameworks* que orientam gestores, educadores e reguladores rumo a uma governança ética, responsável e centrada no bem-estar.

NOTAS

1. Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems, 31*(3), 257-268. https://doi.org/10.1080/0960085x.2022.2026621

2. Duong, C. D., Dao, T. T., Vu, T. N., Ngo, T. V. N., & Tran, Q. Y. (2024, outubro). Compulsive ChatGPT usage, anxiety, burnout, and sleep disturbance: A serial mediation model based on stimulus-organism-response perspective. *Acta Psychologica, 251*, 104622. https://doi.org/10.1016/j.actpsy.2024.104622

3. Preuss, N., Alshehri, A. S., & You, F. (2024, fevereiro). Large language models for life cycle assessments: Opportunities, challenges, and risks. *Journal of Cleaner Production, 466*, 142824. https://doi.org/10.1016/j.jclepro.2024.142824

4. Lu, L., Zhao, J., & Chen, H. (2024, fevereiro). Investigating OTA employees' double-edged perceptions of ChatGPT: The moderating role of organizational support. *International Journal of Hospitality Management, 120*, 103753. https://doi.org/10.1016/j.ijhm.2024.103753

5. Hussain, T., Wang, D., & Li, B. (2024, abril). The influence of the COVID-19 pandemic on the adoption and impact of AI ChatGPT: Challenges, applications, and ethical considerations. *Acta Psychologica, 246*, 104264. https://doi.org/10.1016/j.actpsy.2024.104264

6. Huo, X., & Siau, K. L. (2024). Generative artificial intelligence in business higher education: A focus group study. *Journal of Global Information Management, 32*(1), 1-21. https://doi.org/10.4018/JGIM.364093

7. Lu, L., Zhao, J., & Chen, H. (2024, fevereiro). Investigating OTA employees' double-edged perceptions of ChatGPT: The moderating role of organizational support. *International Journal of Hospitality Management, 120*, 103753. https://doi.org/10.1016/j.ijhm.2024.103753

8. Monib, W. K., Qazi, A., Apong, R. A., Azizan, M. T., Silva, L. De, & Yassin, H. (2024). Generative AI and future education: A review, theoretical validation, and authors' perspective on challenges and solutions. *PeerJ Computer Science, 10*, 1-32. https://doi.org/10.7717/peerj-cs.2105

9. Duong, C. D., Dao, T. T., Vu, T. N., Ngo, T. V. N., & Tran, Q. Y. (2024, outubro). Compulsive ChatGPT usage, anxiety, burnout, and sleep disturbance: A serial mediation model based on stimulus-organism-response perspective. *Acta Psychologica, 251*, 104622. https://doi.org/10.1016/j.actpsy.2024.104622

10. Belanche, D., Belk, R. W., Casaló, L. V., & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal, 44*(3/4), 149-172. https://doi.org/10.1080/02642069.2024.2305451

11. Hussain, T., Wang, D., & Li, B. (2024, abril). The influence of the COVID-19 pandemic on the adoption and impact of AI ChatGPT: Challenges, applications, and ethical considerations. *Acta Psychologica, 246*, 104264. https://doi.org/10.1016/j.actpsy.2024.104264

12. Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society, 75*, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

13. Kumar, S., Rao, P., Singhania, S., Verma, S., & Kheterpal, M. (2024). Will artificial intelligence drive the advancements in higher education? A tri-phased exploration. *Technological Forecasting and Social Change, 201*, 123258. https://doi.org/10.1016/j.techfore.2024.123258

14. Naing, S. Z. S., & Udomwong, P. (2024). Public opinions on ChatGPT: An analysis of Reddit discussions by using sentiment analysis, topic modeling, and SWOT analysis. *Data Intelligence, 6*(2), 344-374. https://doi.org/10.1162/dint_a_00250
15. Saluja, S., Sinha, S., & Goel, S. (2025). Loafing in the era of generative AI. *Organizational Dynamics, 54*(3, Part 2), 101101. https://doi.org/10.1016/j.orgdyn.2024.101101
16. Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805-38818. https://doi.org/10.1109/ACCESS.2023.3268224
17. Hong, S. J. (2025). What drives AI-based risk information-seeking intent? Insufficiency of risk information versus (un)certainty of AI chatbots. *Computers in Human Behavior, 162*, 108460. https://doi.org/10.1016/j.chb.2024.108460
18. Yin, H., Wang, C., & Liu, Z. (2024). Unleashing pre-service language teachers' productivity with generative AI: Emotions, appraisal and coping strategies. *Computers in Human Behavior, 161*, 108417. https://doi.org/10.1016/j.chb.2024.108417
19. Alabed, A., Javornik, A., Gregory-Smith, D., & Casey, R. (2024). More than just a chat: A taxonomy of consumers' relationships with conversational AI agents and their well-being implications. *European Journal of Marketing, 58*(2), 373-409. https://doi.org/10.1108/EJM-01-2023-0037
20. Patvardhan, N., Ranade, M., Vandana, & Khatwani, R. (2024). Examining the accessibility of ChatGPT for competency-based personalised learning for specially-abled people. *International Journal of Information and Learning Technology, 41*(5), 539-555. https://doi.org/10.1108/IJILT-04-2024-0066
21. Milic-Frayling, N. (2023). On the cusp: Computing thrills and perils and professional awakening. *Proceedings of the VLDB Endowment, 16*(12), 4152-4159. https://doi.org/10.14778/3611540.3611640
22. Hao, X., Demir, E., & Eyers, D. (2024). Exploring collaborative decision-making: A quasi-experimental study of human and generative AI interaction. *Technology in Society, 78*, 102662. https://doi.org/10.1016/j.techsoc.2024.102662
23. Monib, W. K., Qazi, A., Apong, R. A., Azizan, M. T., Silva, L. De, & Yassin, H. (2024). Generative AI and future education: A review, theoretical validation, and authors' perspective on challenges and solutions. *PeerJ Computer Science, 10*, 1-32. https://doi.org/10.7717/peerj-cs.2105
24. Lian, Y., Tang, H., Xiang, M., & Dong, X. (2024). Public attitudes and sentiments toward ChatGPT in China: A text mining analysis based on social media. *Technology in Society, 76*, 102442. https://doi.org/10.1016/j.techsoc.2023.102442
25. Baldry, M. K., Happa, J., Steed, A., Smith, S., & Glencross, M. (2024). From embodied abuse to mass disruption: Generative, inter-reality threats in social, mixed-reality platforms. *Digital Threats: Research and Practice, 5*(4), Article 38, 1–36. https://doi.org/10.1145/3696015
26. Karagianni, A., & Doh, M. (2024). A feminist legal analysis of non-consensual sexualized deepfakes: Contextualizing its impact as AI-generated image-based violence under EU law. *Porn Studies,* v, n? 1-18. https://doi.org/10.1080/23268743.2024.2408277
27. Ozuem, W., Ranfagni, S., Willis, M., Salvietti, G., & Howell, K. (2024). Exploring the relationship between chatbots, service failure recovery and customer loyalty: A frustration–aggression perspective. *Psychology & Marketing, 41*(10), 2253-2273. https://doi.org/10.1002/mar.22051
28. Brüns, J. D., & Meißner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services, 79*, 103790. https://doi.org/10.1016/j.jretconser.2024.103790
29. Trincado-Munoz, F. J., Cordasco, C., & Vorley, T. (2024). The dark side of AI in professional services. *The Service Industries Journal, 45*(5/6), 455-474. https://doi.org/10.1080/02642069.2024.2336208
30. Belanche, D., Belk, R. W., Casaló, L. V., & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal, 44*(3/4), 149-172. https://doi.org/10.1080/02642069.2024.2305451
31. Holm, J. R., Hain, D. S., Jurowetzki, R., & Lorenz, E. (2023). Innovation dynamics in the age of artificial intelligence: Introduction to the special issue. *Industry and Innovation, 30*(9), 1141-1155. https://doi.org/10.1080/13662716.2023.2272724
32. Cao, J., & Song, Z. (2025). An incoming threat: The influence of automation potential on job insecurity. *Asia-Pacific Journal of Business Administration, 17*(1), 116-135. https://doi.org/10.1108/APJBA-07-2022-0328
33. Wei, X., Chu, X., Geng, J., Wang, Y., Wang, P., Wang, H. X., Wang, C., & Lei, L. (2024). Societal impacts of chatbot and mitigation strategies for negative impacts: A large-scale qualitative survey of ChatGPT users. *Technology in Society, 77*(59), 102566. https://doi.org/10.1016/j.techsoc.2024.102566
34. Gandía, J. A. G., Gavrila, S. G., Ancillo, A. de L., & Núñez, M. T. del V. (2025). Towards sustainable business in the automation era: Exploring its transformative impact from top management and employee perspective. *Technological Forecasting and Social Change, 210,* 123908. https://doi.org/10.1016/j.techfore.2024.123908
35. Belanche, D., Belk, R. W., Casaló, L. V., & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal, 44*(3/4), 149-172. https://doi.org/10.1080/02642069.2024.2305451
36. Gupta, R., & Rathore, B. (2024, julho). Exploring the generative AI adoption in service industry: A mixed-method analysis. *Journal of Retailing and Consumer Services, 81*, 103997. https://doi.org/10.1016/j.jretconser.2024.103997
37. Kiuchi, K., Otsu, K., & Hayashi, Y. (2023). Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: A systematic meta-review. *Behaviour and Information Technology, 43*(15), 3696-3736. https://doi.org/10.1080/0144929X.2023.2286528
38. French, A. M., & Shim, J. P. (2024). From artificial intelligence to augmented intelligence: A shift in perspective, application, and conceptualization of AI. *Information Systems Frontiers. 27*(4), 1345-1366. https://doi.org/10.1007/s10796-024-10562-2
39. Wei, X., Chu, X., Geng, J., Wang, Y., Wang, P., Wang, H. X., Wang, C., & Lei, L. (2024). Societal impacts of chatbot and mitigation strategies for negative impacts: A large-scale qualitative survey of ChatGPT users. *Technology in Society, 77*(59), 102566. https://doi.org/10.1016/j.techsoc.2024.102566
40. Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour and Information Technology, 44*(11), 2518-2544. https://doi.org/10.1080/0144929X.2024.2394886

41. Stahl, B. C., & Eke, D. (2024, abril). The ethics of ChatGPT: Exploring the ethical issues of an emerging technology. *International Journal of Information Management, 74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700
42. Cao, J., & Song, Z. (2025). An incoming threat: The influence of automation potential on job insecurity. *Asia-Pacific Journal of Business Administration, 17*(1), 116-135. https://doi.org/10.1108/APJBA-07-2022-0328
43. Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., Santini, F. de O., Ladeira, W. J., Rather, R. A., Yasin, N., Rodriguez, R. V., Kokkalis, P., Murad, M. W., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era: A holistic framework. *International Journal of Management Education, 22*(3), 101041. https://doi.org/10.1016/j.ijme.2024.101041
44. Duke Learning Innovation & Lifetime Education. (2025, agosto 25). *Artificial intelligence policies: Guidelines and considerations.* https://lile.duke.edu/ai-and-teaching-at-duke-2/artificial-intelligence-policies-in-syllabi-guidelines-and-considerations/; Fish, L. (2025, setembro 10). *Artificial intelligence and mental health: 6 essential ways to protect your mind while engaging with AI.* https://laura-fishtherapy.com/artificial-intelligence-mental-health-ways-to-protect-your-mind/; Kaul, A., Ng, M., Lim, M., Goh, L., & Brahmecha, A. (2025, fevereiro 5). *12 key principles for sustainable AI.* Access partnership. https://accesspartnership.com/opinion/12-key-principles-for-sustainable-ai; Loh, H., Beauchene, V., Lukic, V., & Shenov, R. (2024, outubro 8). *Five must-haves for effective AI upskilling. BCG.* https://www.bcg.com/publications/2024/five-must-haves-for-ai-upskilling; Nagenda, M. (2023, setembro 8). *Increasing awareness about the ethical implications of AI.* Pranathi Software Services Pvt. Ltd. https://www.pranathiss.com/blog/the-ethical-implications-of-ai/
45. Demetis, D. S., & Angell, I. O. (2024, fevereiro). *The technological construction of reality* (Vol. 2). Edward Elgar Publishing Limited.

NOTA

A ferramenta SciSpace foi utilizada e citada na metodologia; o ChatGPT e o Google Tradutor auxiliaram na tradução de alguns termos, e o ChatGPT também auxiliou na verificação da ortografia e da gramática em trechos pontuais. Após utilizar essas ferramentas, revisamos e editamos o conteúdo conforme necessário, assumindo total responsabilidade por ele.

CONFLITOS DE INTERESSE

As autoras não têm conflito de interesse a declarar.

MINICURRÍCULO

Suely Fischer Omura é professora na FGV-EAESP, doutora em Administração de Empresas pela mesma instituição e pesquisa inteligência artificial, sistemas de informação e processo decisório, com foco na interação humano–IA.
Vanessa Martins dos Santos é professora na ESPM Escola Superior de Propaganda e Marketing e na FIA Fundação Instituto de Administração, doutora em Administração de Empresas pela FGV-EAESP e pesquisa inteligência artificial, blockchain e sustentabilidade.

CONTRIBUIÇÃO DE AUTORIA

Suely Fischer Omura: Conceitualização, curadoria de dados, análise formal; Metodologia; Redação – rascunho original.
Vanessa Martins dos Santos: Conceitualização, curadoria de dados, análise formal; Metodologia; Redação – rascunho original.

Submetido 14-10-2025. **Aprovado** 03-02-2026
Avaliado pelo sistema de revisão duplo-anônimo.
Primeiro(a) avaliador(a): não autorizou a divulgação de sua identidade e relatório de avaliação por pares.
Segundo(a) avaliador(a): Alexandre Cappellozza e, não autorizou a divulgação de sua identidade.
Versão original. DOI: http://dx.doi.org/10.12660/gvexec.v25n4.2026.94510