LUMEN ET VIRTUS
ISSN: 2177-2789

IMPACTOS EMOCIONAIS E AFETIVOS DO USO DA INTELIGÊNCIA ARTIFICIAL

https://doi.org/10.56238/levv16n44-026

Data de submissão: 13/12/2024 **Data de publicação:** 13/01/2025

Flávia Ladeira Lippi
Instituto de Desenvolvimento Humano Lippi
Mestranda em Tecnologias da Inteligência e Design Digital

Carolina Cássia Conceição Abilio
Instituto de Desenvolvimento Humano Lippi
Mestre em Ciências, Universidade de São Paulo

José Raimundo da Silva Lippi
Instituto de Desenvolvimento Humano Lippi
Doutor em Saúde da Criança e da Mulher

Daniel Couto Gatti
Pontifícia Universidade Católica de São Paulo
Doutor em Educação Matemática

Marcelo Augusto Vieira Graglia
Pontifícia Universidade Católica de São Paulo
Doutorado em Tecnologias da Inteligência e Design Digital

RESUMO

A crescente adoção de tecnologias baseadas em Inteligência Artificial (IA) tem gerado significativo interesse social e acadêmico, especialmente com a rápida adoção de sistemas de IA generativa, como o ChatGPT. Essas ferramentas, projetadas para simular a comunicação humana, tornaram-se onipresentes, transformando interações individuais, organizacionais e sociais. Este artigo explora a interseção entre a IA e as experiências emocionais e afetivas humanas, com foco nos impactos da IA generativa na saúde mental, nos relacionamentos interpessoais e nos comportamentos sociais. Com base nos avanços recentes em IA, o estudo apresenta uma análise crítica dos benefícios e riscos associados a essas tecnologias, incluindo seu potencial de democratização e acessibilidade, bem como os desafios éticos representados pelos vieses embutidos em seus algoritmos. A análise destaca também casos ilustrativos em que a IA generativa influenciou o bem-estar emocional, enfatizando o papel dual dessas tecnologias como facilitadoras de conexão e potenciais fontes de dano. O artigo conclui com a importância das discussões sobre a ética no desenvolvimento e uso da inteligência artificial e sobre as questões de governança e regulação como garantias para o desenvolvimento e para a aplicação responsável das tecnologias de IA em contextos diversos.

Palavras-chave: Inteligência Artificial. Inteligência Artificial Generativa. Saúde Mental. Computação Afetiva. Ética em IA.

1 INTRODUÇÃO

1.1 HYPE DE IA VERSUS A REALIDADE

Segundo LaGrandeur (2023), estamos vivendo um momento de *hype* de soluções tecnológicas baseadas em Inteligência Artificial (IA). Embora essas tecnologias possam potencialmente impactar de maneira positiva diversas áreas da sociedade, seu uso sem crivo crítico e entendimento de seu potencial pode trazer mais malefícios do que benefícios. Nesse sentido, o *hype* em torno da IA aumenta a possibilidade de consequências negativas para a sociedade, incluindo o campo da segurança pública, dos direitos humanos e os ambientes empresarial e educacional (GRAGLIA *et al.*, 2024).

Em 2022 cerca de 50% das empresas globais já haviam adotado processos de IA em ao menos uma área interna e 44% das empresas já buscavam implementar IA em seus processos e serviços existentes. Em 2023 o uso de IA cresceu 250% (MELLO, 2024). Uma das aplicações mais buscadas tem sido a utilização para automatizar tarefas em larga escala ou ampliar a capacidade de análise dos dados disponíveis nas bases das organizações. Nestes casos, as tecnologias baseadas em IA podem contribuir para a melhoria do desempenho dos processos e serviços corporativos, gerando impactos positivos nos principais indicadores de eficiência do negócio. A previsão é de um mercado global de produtos baseados em IA de mais de 100 bilhões de dólares até 2028 (MAHESHWARI, 2024; ENHOLM et al, 2021).

As previsões foram afetadas pelo surgimento dos chamados modelos de linguagem de grande escala – *large Language models* ou *LLM* – que são treinados a partir de enormes quantidades de dados não rotulados usando técnicas de aprendizado de máquina não supervisionado (NAVEED *et al*., 2024).

Estes modelos, classificados como de propósito geral, são compostos por redes neurais artificiais com bilhões de parâmetros e apresentam alto desempenho em muitos tipos de aplicação, especialmente aquelas que envolvem o campo da linguagem natural. As redes neurais constituem uma subcategoria do aprendizado de máquina (*machine learning*), caracterizada pela capacidade dos sistemas de aprender com erros e aprimorar seu desempenho de forma contínua. (AMAZON, 2024).

Tais modelos também são chamados de IA generativa, por conta de sua capacidade de gerar textos, desenhos, músicas, códigos, entre outros, a partir dos dados disponíveis nas bases existentes. O ChatGPT, um produto desenvolvido pela empresa estadunidense OpenAI e lançado no mercado em novembro de 2022, destacou-se como uma das primeiras inteligências artificiais generativas amplamente acessíveis ao público geral. Isto foi significativo, dado que grande parte das soluções em IA tinham como foco exclusivamente o mercado corporativo e seu manuseio era restrito basicamente a pesquisadores e especialistas. Em janeiro de 2023, a plataforma atingiu a marca de 100 milhões de usuários globalmente, um recorde em termos de velocidade de expansão entre usuários no cenário digital (SINGH, 2023).

Além do ChatGPT, outros modelos baseados em arquiteturas de aprendizado profundo têm se popularizado na forma de produtos comerciais, como o Google Gemini e o Microsoft Copilot. Tecnologias similares também têm sido desenvolvidas em contextos não ocidentais, como o ERNIE Bot, na China, e o YaLM, na Rússia (HU, 2023; SINGH, 2023).

Embora o surgimento de tecnologias baseadas em IA não seja recente, considerando que a área de estudos dedicada a esta temática vem sendo desenvolvida há décadas, pode-se afirmar que a inovação fundamental promovida pela OpenAI e outras empresas do setor, além das características técnicas que definem a chamada inteligência artificial generativa, foi a capacidade de tornar uma tecnologia altamente complexa acessível e simples de usar. Ao desenvolver ferramentas que não demandam conhecimentos técnicos avançados por parte dos usuários, essas organizações popularizaram o acesso à tecnologia, expandindo o uso da IA para além do domínio de empresas, universidades e especialistas, inserindo esta tecnologia no cotidiano de milhões de pessoas e consolidando-a como uma ferramenta de uso comum e intuitivo em diversas atividades diárias (MINH, 2021).

À despeito das potencialidades positivas das ferramentas de IA generativa (MCKINSEY & COMPANY, 2024), há aspectos que envolvem riscos, especialmente aqueles derivados dos vieses que podem ser incorporados pelos algoritmos a partir do treinamento e uso de bases de dados. Os resultados podem ser classificações, recomendações, ou outras saídas geradas pelos algoritmos, que reflitam padrões de discriminação e exclusão que impactem negativamente determinados grupos sociais, refletindo desigualdades estruturais de longa data em nossa sociedade (AKTER *et al*., 2021) e mesmo desvios éticos.

Este artigo tem como objetivo discutir os impactos de sistemas baseados em inteligência artificial – mais especificamente, a IA generativa – nas emoções e afetos humanos. Para isso, baseou-se numa revisão da bibliografia especializada e em casos ilustrativos de como sistemas de IA estão impactando aspectos emocionais e a saúde mental de indivíduos em diferentes contextos.

2 IMPACTOS DA TECNOLOGIA NA EMOÇÃO HUMANA

Existem impactos emocionais e afetivos relacionados ao uso da tecnologia e, mais especificamente, da inteligência artificial. Para se aprofundar na análise das relações complexas entre seres humanos e tecnologia, foi proposto, em 1997, o conceito de computação afetiva (PICARD, 1997).

Essa é uma área acadêmica que abrangente emoções humanas, sentimentos e percepções emergentes na interface com alguma tecnologia. Mais recentemente, expandiu-se para também englobar o reconhecimento de emoções e análise de sentimentos por parte de soluções tecnológicas, incluindo, dentro elas, as ferramentas baseadas em IA. Conceitua-se, a partir da computação afetiva,

que o entendimento das emoções humanas na interface com a tecnológica é fundamental para desenvolver e ampliar a efetividade de soluções baseadas em IA centradas na inteligência humana (WANG *et al*., 2022).

Tal como qualquer tecnologia, a IA não opera em um vácuo, mas sim em um contexto social, cultural, econômico e geopolítico complexo, e os impactos emocionais e afetivos podem assumir tanto características positivas quanto negativas, seja por omissão de seus criadores ou de maneira intencional. No contexto brasileiro, um exemplo notável de impacto afetivo de grande alcance foi a campanha publicitária realizada patrocinada pela Volkswagen em 2023 para comemorar os 70 anos da empresa. Em um vídeo, a imagem da cantora Elis Regina, uma das artistas mais populares da Música Popular Brasileira, é recriada através de tecnologia *deepfake* para que parecesse cantando ao lado da filha, Maria Rita. Um dueto real nunca foi possível por conta do falecimento de Elis em 1982, aos 36 anos, quando Maria Rita tinha apenas 4 anos (BISCHOFF, 2023).

A campanha é considerada um marco do uso de IA na publicidade brasileira, sendo a primeira exposição ampla deste tipo de tecnologia para grande parte da população. Com isso, iniciou-se um debate na esfera pública acerca dos limites da IA e os aspectos éticos relacionados a essa tecnologia. Rosa (2024), em uma pesquisa qualitativa analisando as percepções do público sobre a publicidade, afirma que "o comercial trouxe mais simpatia, nostalgia e carinho. Em outras palavras, construiu memória positiva de marca" (ROSA, 2024, pg. 68).

Porém, questões éticas relacionadas ao direito de imagem, disseminação de desinformação, manipulação e falta de compromisso com a verdade também foram aspectos abordados pelo público, que, de modo geral, espera que as novas tecnologias devam ser utilizadas pelas marcas e empresas "na medida certa" e sua utilização deva ser guiada por princípios éticos e causar impactos positivos na humanidade, servindo genuinamente para ajudar as pessoas (ROSA, 2024).

3 IMPACTOS NEGATIVOS DA IA GENERATIVA NAS EMOÇÕES

A capacidade de criar entrevistas e vídeos falsificados, nos quais uma pessoa aparece dando declarações que nunca realizou, já é uma realidade tecnológica com potencial para enganar e manipular outras pessoas, principalmente em contextos políticos. Esses conteúdos são criados através de tecnologia generativa de inteligência artificial e são denominados *deepfakes* (FARID,2019).

Um exemplo notável ocorreu nos Estados Unidos, em 2023, quando circulou nas redes sociais um vídeo manipulado do então presidente estadunidense Joe Biden fazendo declarações transfóbicas e críticas negativas a grupos sociais minoritários. Embora não tenha sido um caso isolado, o acontecimento tornou-se ilustrativo dos riscos associados ao uso de tecnologias avançadas de geração de mídia sintética, especialmente em contextos políticos e sociais impetuosos (REUTERS FACT CHECK, 2023).

Após gerar controvérsias significativas junto à opinião pública nos Estados Unidos, uma investigação conduzida por um veículo de imprensa revelou que o vídeo havia sido criado por um usuário do Twitter utilizando tecnologia *deepfake*. Porém, até que essa informação fosse divulgada, o impacto do conteúdo já havia se consolidado, especialmente em um contexto em que grande parte da população ainda não estava familiarizada com a existência de tecnologias de manipulação audiovisual com características tão realistas. A utilização deste tipo de recurso pode gerar consequências negativas para a coesão social e para a estabilidade da democracia.

Segundo Samer e outros (2023), os *deepfakes* podem servir para influenciar significativamente as escolhas pessoais de indivíduos e a sociedade como um todo. Uma das consequências mais profundas está relacionada à crise de desinformação e a quebra de confiança nas instituições. Neste contexto, torna-se difícil separar o falso do verdadeiro, culminando em sociedades pautadas na pós-verdade onde a confiança entre pessoas e instituições é corroída por conteúdos falsos (CHESNEY, CITRON, 2019).

Os impactos causados pela tecnologia *deepfake* não se restringem a figuras públicas ou de importância política. Um relatório publicado em 2023 revelou que o número de vídeos criados com tecnologia *deepfake* na internet aumentou 550% desde 2019. Ao todo, 98% desses vídeos possuem conteúdo de natureza pornográfica, e 99% das pessoas retratadas nesses materiais falsificados são mulheres. Este cenário destaca uma tendência alarmante de uso da tecnologia para fins prejudiciais e maliciosos, expondo questões éticas e de privacidade, sobretudo no que diz respeito à vulnerabilidade de mulheres em ambientes digitais (SECURITY HERO, 2023).

A utilização de *deepfakes* para criação de conteúdo explícito, manipulação de pessoas e disseminação de informações falsas corrói a confiança individual e social, criando consequências negativas a curto, médio e longo prazo. Esse cenário torna-se ainda mais grave ao considerar que os avanços recentes em tecnologias generativas de efeitos visuais e sonoros hiper-realistas podem, em breve, tornas vídeos *deepfake* indistinguíveis de vídeos autênticos (GROH *et al*., 2024).

Tais riscos são particularmente preocupantes no cenário brasileiro. Um estudo realizado pela Organização para a Cooperação e Desenvolvimento Econômico (2024) avaliou a capacidade de identificação de notícias falsas ou enganosas entre voluntários de 21 países, e mostrou que o Brasil apresenta os menores índices de consciência emocional e de discernimento relacionados ao uso de tecnologias baseadas em IA. Além disso, a pesquisa revelou uma correlação importante entre o uso intensivo de redes sociais como principal fonte de informação e a maior propensão a acreditar em notícias falsas, destacando a vulnerabilidade da população brasileira em relação à desinformação gerada por *deepfakes* (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2024).

É também fundamental discutir o impacto emocional negativo sob a perspectiva de gênero, considerando que mulheres vítimas dessa prática têm suas identidades apropriadas para fins com os quais não consentem. A utilização da IA como ferramenta, nesse contexto, reflete uma cultura estruturalmente machista que instrumentaliza o corpo feminino para atender aos interesses de determinados grupos. Laffier e Rehman (2023) mostram como mulheres alvo de *deepfakes* sofrem consequências emocionais e sociais a longo prazo, como sentimentos de medo, desajustes psicológicos e sociais, transtornos de saúde mental e danos à sua reputação profissional. Isso impacta diretamente suas relações sociais e profissionais, causando até mesmo prejuízos econômicos.

No campo do trabalho, segundo Cremer e Koopman (2024), o crescimento do uso de IA está associado a impactos negativos no bem-estar e no engajamento dos colaboradores. Um levantamento realizado com profissionais de Taiwan, Indonésia, Estados Unidos e Malásia demonstrou que, embora a IA esteja permitindo aos trabalhadores maiores índices de produtividade, essa eficiência vem acompanhada de uma redução no tempo dedicado à interação com colegas. Consequentemente, surgem perdas significativas em oportunidades de colaboração e conectividade dentro das equipes, comprometendo aspectos essenciais do trabalho em grupo (TANG, KOOPMAN, MAI, CREMER, ZHANG, REYNDERS *et al*, 2023).

A eficiência organizacional alcançada com o uso de tecnologias baseadas em IA, quando estas não são implementadas de modo a complementar a cultura organizacional, geram um custo elevado em termos de bem-estar dos profissionais. O uso excessivo e inadequado da IA em algumas empresas aumenta o isolamento social entre os colaboradores, impactando negativamente a qualidade de vida. Os colaboradores relatam insônia, abuso de álcool, piora no humor, declínio na função cognitiva e o surgimento de problemas de saúde física e mental (TANG, KOOPMAN, MAI, CREMER, ZHANG, REYNDERS *et al*., 2023).

As mudanças nas dimensões do ambiente de trabalho e as relações socioemocionais profissionais são debatidas por Graglia e Lazzareschi (2018, 2023) no contexto de introdução de novas tecnologia em contextos organizacionais. Nessa perspectiva, pode-se observar a importância de os trabalhadores preservarem e aprimorarem uma visão sistêmica de suas atividades profissionais, integrando as novas tecnologias como ferramentas de suporte ao desempenho laboral, e não como um objetivo final em si mesmas. De qualquer forma, o avanço dos processos de transformação digital tem impactado de forma severa os modos e as relações de trabalho (GRAGLIA; LAZZARESCHI; HUELSEN, 2021).

Um outro aspecto que envolve impactos emocionais na relação individual com as tecnologias baseadas em IA é a ocorrência cada vez mais frequente do uso de ferramentas de IA para interação em momentos de solidão ou como forma de desabafo. Embora existam chatbots desenvolvidos especificamente para fins terapêuticos, com apoio de profissionais de saúde mental e treinamento

direcionado a esse propósito, este tipo de uso é normalmente realizado com aplicações mais populares e de fácil acesso, como o ChatGPT. Acontece, que essa tecnologia específica foi projetada para desempenhar outras funções que não incluem abordar adequadamente questões emocionais dos usuários, gerando impactos emocionais e afetivos negativos (AMARAL, MEIRELES, 2023).

Em 2024, um exemplo desta situação ganhou notoriedade global com o suicídio de um jovem de 14 anos na Flórida, Estados Unidos, após este se envolver em um relacionamento com um chatbot da empresa Character.IA, uma plataforma com mais de 20 milhões de usuários, criada pela Google. A mãe do jovem abriu um processo contra a empresa responsável por essa tecnologia, alegando que a organização criou, ou facilitou, um cenário de dependência emocional com o seu produto. Segundo o processo, ocorreram "interações abusivas e de cunho sexual" ao longo de um período de dez meses entre o chatbot e o jovem, culminando em uma relação de abuso emocional (HETZNER, 2024).

A Character.IA, assim como outras empresas que oferecem soluções similares, é uma plataforma onde os chamados “companheiros de IA” simulam relacionamentos afetivos com os usuários, podendo ser de natureza romântica ou sexual. Um documento elaborado em parceria com profissionais de saúde mental do Stanford Brainstorm Lab define essa tecnologia da seguinte maneira:

> (...) os companheiros de IA representam “uma nova categoria de tecnologia que vai além dos chatbots simples”. Eles são projetados, entre outras funcionalidades, para “simular vínculos emocionais e relacionamentos próximos com os usuários, lembrar detalhes pessoais de conversas anteriores, desempenhar papéis como mentores e amigos, imitar emoções humanas e empatia, e concordar mais prontamente com o usuário do que os chatbots de IA típicos” (GREENFIELD, 2024; COMMON SENSE MEDIA, 2024).

O ex-CEO da Google, Eric Schmidt, comentou em uma entrevista suas preocupações de que a IA possa, em breve, ser capaz de oferecer parceiros românticos ideais para adolescentes e jovens adultos que têm dificuldade em se relacionar com outras pessoas. Segundo ele, a indústria de tecnologia como um todo otimiza seus produtos para maximizar a atenção de seus usuários e monetizá-la. Isso pode levar o mercado de IA a criar soluções cada vez mais personalizadas para causar dependência emocional em seus usuários caso não haja regulações éticas que possam frear esse crescimento (HETZNER, 2024).

Há questões éticas significativas envolvidas neste caso. Por exemplo, a Character.IA, mesmo tendo ciência dos pensamentos suicidas expressadas por seu usuário, falhou em notificar seus responsáveis ou as autoridades locais. Modelos de IA não treinados especificamente com dados voltados para acolher e apoiar a saúde mental de seus usuários podem apresentar riscos ainda mais significativos a pessoas vulneráveis. Isso pode acarretar consequências adversas para a saúde mental, bem como para diversos outros aspectos da vida do indivíduo, devido ao suporte inadequado – e, muitas vezes, prejudicial – fornecido em contextos emocionalmente sensíveis.

4 IMPACTOS POSITIVOS DA IA GENERATIVA NAS EMOÇÕES

Com os avanços recentes no campo da inteligência artificial generativa tem surgido uma "classe de tecnologias de computação afetiva utilizada na computação pessoal e social, englobando técnicas emergentes e controversas que visam classificar emoções humanas e outros fenômenos afetivos" (NAZANIN ANDALIBI *et al*., 2024; tradução dos autores).

Embora ainda haja controvérsia em termos dos benefícios e malefícios desta tecnologia nas aplicações relacionadas ao suporte psicológico, como violações de privacidade e a forma como estes produtos têm sido comercializados, algumas soluções já vêm sendo utilizadas pelo público em geral com resultados positivos, indo além de efeitos indiretos proporcionados por conteúdos como vídeos ou imagens. Sistemas de IA baseados em *feedback* em tempo real têm a capacidade de ajustar seu comportamento com base nas respostas dos usuários, promovendo melhorias na experiência interativa para alcançar objetivos pré-definidos.

Aplicações práticas dessas tecnologias demonstram resultados promissores em diferentes áreas, como no campo da saúde mental e educação. Neste contexto, pesquisadores finlandeses estão desenvolvendo um sistema capaz de identificar níveis de estresse a partir dos padrões de uso do mouse em computadores. Essas informações geram alertas e direcionam os indivíduos para intervenções adequadas antes que o estresse atinja níveis mais graves. Já na área da educação, sistemas de tutoria inteligentes estão sendo treinados para oferecem experiências de aprendizado personalizadas, permitindo que os alunos estudem em seu próprio ritmo e recebam *feedback* customizado com base na interação do usuário com o sistema. Essa abordagem melhora os resultados de aprendizagem e permite os alunos a personalizarem sua experiência de aprendizagem (MUSTAFA, 2024).

No campo da saúde mental, destacam-se os aplicativos de apoio emocional, projetados para oferecer suporte terapêutico em casos de dificuldade de acesso do usuário a profissionais de saúde mental, como é o caso do Woebot, um chatbot de saúde mental baseado em IA capaz de simular conversas, interpretar o que os usuários comunicam por meio de mensagens escritas ou faladas e oferecer respostas adequadas às interações (DARCY *et al.*, 2019).

Este sistema foi treinado com o auxílio de profissionais de saúde mental para fornecer suporte emocional e terapêutico aos usuários para reduzir sintomas de ansiedade e depressão e, para isso, usa os dados gerados pelas nas interações realizadas e as informações fornecidas por cada usuário (DURDEN *et al*., 2023).

Pesquisas indicam que o uso do aplicativo esteve associado a reduções significativas de estresse e *burnout* nos usuários, além de melhora dos índices associados à resiliência em um período de oito semanas, especialmente em usuários com sintomas piores de ansiedade ou depressão. Embora pontuais, esses resultados ilustram como soluções baseadas em IA têm o potencial de fornecer suporte a pessoas com sintomas clínicos que enfrentam barreiras no acesso a profissionais de saúde mental

(DURDEN *et al*., 2023), desde que não sejam encaradas como um recurso que substitua os tratamentos conduzidos por profissionais especializados. As aplicações de IA voltadas para a saúde mental podem ser amplamente escaláveis, desempenhando um papel importante na pré-triagem de pacientes e no manejo de transtornos mentais leves, ampliando de maneira significativa o alcance dos cuidados psicológicos (LIPPI, ABILIO, LIPPI, GRAGLIA, 2024).

5 ÉTICA, GOVERNANÇA E REGULAÇÃO DA IA

Segundo Crawford (2021), a IA é, em sua essência, uma tecnologia de extração, cuja lógica se manifesta desde a obtenção dos minerais utilizados em sua infraestrutura até a coleta de dados, muitas vezes realizada sem o consentimento dos indivíduos. Esse processo inclui a captura de ações e expressões coletivas, consolidando um modelo que privilegia a exploração em detrimento da reciprocidade, com consequências possivelmente graves, a depender da natureza do seu uso.

Contrariamente ao ideal de uma via de mão dupla, como proposto por pesquisadores que promovem a chamada Inteligência Colaborativa, a IA frequentemente opera como um sistema que avança de forma descontrolada sobre os direitos e a privacidade das pessoas, exacerbando desigualdades estruturais. Este cenário está conduzindo a um modelo de governança caracterizado por tendências antidemocráticas e ao agravamento das desigualdades sociais, uma vez que concentra o poder e o conhecimento sobre tecnologias baseadas em IA em um número restrito de grandes corporações.

Para mitigar esses riscos, torna-se imprescindível debater e implementar regulamentações e diretrizes éticas que orientem tanto o desenvolvimento como a aplicação da IA. Um exemplo significativo dessa iniciativa é observado na União Europeia, que tem liderado esforços nesse sentido (EUROPEAN COMMISSION, 2021; AGÊNCIA BRASIL, 2024).

No entanto, destaca-se o quão recente é esse movimento: a primeira legislação mundial voltada especificamente para a regulamentação da IA foi publicada apenas em 2024 e entrará em vigor em 2026 (CHEE; HUMMEL, 2024; CRAWFORD, 2021). No Brasil, as discussões sobre os riscos perpassam o direito à informação, quando sistemas de IA utilizam dados pessoais, o direito à privacidade, o direito a não discriminação, e a necessidade de se utilizar linguagem clara e acessível ao público leigo (SENADO FEDERAL, 2024).

Embora esses debates representem passos iniciais, é essencial ampliar e fortalecer espaços de discussão, tanto em arenas formais como informais, considerando que a IA não é uma tecnologia isolada, mas sim um sistema com impacto abrangente a nível individual, organizacional e coletivo. Portanto, o impacto da sua utilização vai além dos seus usuários diretos, afetando também os indivíduos cujos dados são coletados e utilizados para o treinamento dos modelos algorítmicos.

Privacidade, ética e consentimento emergem como elementos centrais no debate sobre os potenciais impactos negativos do uso de tecnologias de IA em larga escala, principalmente considerando-se os elementos emocionais e afetivos associados à essas tecnologias, de maneira intencional ou não. Esses aspectos tornam-se particularmente relevantes ao observamos as implicações a longo prazo associadas à utilização indiscriminada dessas tecnologias, que podem envolver a perpetuação de estereótipos e exclusão de grupos minoritários, a amplificação de preconceitos e o agravamento das desigualdades sociais, a corrosão da confiança pública nas organizações e a intensificação de divisões estruturais.

Em 2021, a Organização Mundial da Saúde publicou em seu relatório *Ethics & Governance of Artificial Intelligence for Health* seis princípios éticos que orientam o uso responsável da IA no campo da saúde (WORLD HEALTH ORGANIZATION, 2021). Estes pilares norteadores apresentam relevância transversal e podem ser contemplados por organizações que desejam empregar soluções baseadas em IA de maneira ética. Também podem pautar os debates públicos para o desenvolvimento de regulamentações e sistemas de governança para garantir o bem-estar das pessoas e os interesses coletivos. Os princípios incluem:

i. Proteger a autonomia das pessoas;
ii. Promover o bem-estar humano, a segurança humana e o interesse público;
iii. Garantir transparência, explicabilidade e inteligibilidade dos sistemas de IA;
iv. Fomentar a responsabilidade e a prestação de contas;
v. Garantir inclusão e equidade na sociedade;
vi. Promover uma IA responsiva e sustentável.

6 CONSIDERAÇÕES FINAIS

As soluções baseadas em IA têm o potencial de impactar positivamente ou negativamente os indivíduos, as organizações e a sociedade, conforme a aplicação e a forma como as soluções são desenvolvidas. Especificamente, é necessário aprofundar as discussões relacionadas às questões éticas do uso de tecnologias de inteligência artificial relacionadas às emoções de seus usuários, principalmente considerando-se o contexto comercial e a falta de uma regulação governamental adequada para garantir o bem-estar dos seus usuários.

Um dos principais desafios está justamente na falta de transparência dos modelos de IA em relação aos critérios usados para fundamentar suas decisões. Essa opacidade pode tanto levar à marginalização de grupos sociais minoritários e intensificar a exclusão social dessas populações, como ocultar riscos na interação de usuários com estes sistemas. Isso porque a ausência de clareza sobre os processos decisórios dificulta a identificação de possíveis vieses e, consequentemente, afasta a

implementação de medidas para mitigá-los, perpetuando desigualdades sociais e limitando a possibilidade de intervenções corretivas (AI NOW INSTITUTE, 2024).

Embora a implementação de princípios éticos seja desafiadora, especialmente considerando as transformações profundas decorrentes do uso recente das tecnologias baseadas em IA, os avanços registrados na literatura sugerem que é possível direcionar o uso dessas inovações para fins que impactem positivamente a sociedade, alinhando a evolução tecnológica com os direitos fundamentais individuais e coletivos.

REFERÊNCIAS

AI NOW INSTITUTE. About Us. [~2024]. Disponível em: https://ainowinstitute.org/about. Acesso em: 29 jul. 2024.

AKTER, S. et al. Algorithmic bias in data-driven innovation in the age of AI. International Journal of Information Management, v. 60, p. 102387, out. 2021.

AMARAL, R.; MEIRELES, C. ChatGPT no lugar de terapia? Psicóloga lista os perigos dessa prática. Metrópoles, 2023. Disponível em: https://www.metropoles.com/colunas/claudia-meireles/chatgpt-no-lugar-de-terapia-psicologa-lista-os-perigos-dessa-pratica. Acesso em: 29 jul. 2024.

AMAZON WEB SERVICES, INC. O que é uma rede neural? Guia de IA e ML - AWS. [~2024]. Disponível em: https://aws.amazon.com/pt/what-is/neural-network/. Acesso em: 29 jul. 2024.

BISCHOFF, W. Elis Regina aparece cantando ao lado da filha Maria Rita em campanha feita com inteligência artificial. São Paulo: G1, 2023. Disponível em: https://g1.globo.com/economia/midia-e-marketing/noticia/2023/07/04/elis-regina-aparece-cantando-ao-lado-da-filha-maria-rita-em-campanha-da-volkswagen-feita-com-inteligencia-artificial.ghtml. Acesso em: 29 jul. 2024.

CHEE, F. Y.; HUMMEL, T. Europa cria referência global, com legislação para regulamentar IA. Agência Brasil, 2024. Disponível em: https://agenciabrasil.ebc.com.br/internacional/noticia/2024-05/europa-cria-referencia-global-com-legislacao-historica-para-conter-ia. Acesso em: 29 jul. 2024.

CHESNEY, R.; CITRON, D. Deepfakes and the New Disinformation War: the coming age of post-truth geopolitics. Foreign Affairs, v. 98, p. 147, 2019.

COMMON SENSE MEDIA. Parents' Ultimate Guide to AI Companions and Relationships. 2024. Disponível em: https://www.commonsensemedia.org/articles/parents-ultimate-guide-to-ai-companions-and-relationships. Acesso em: 1 dez. 2024.

CRAWFORD, K. The Atlas of AI: power, politics, and the planetary costs of artificial intelligence. Dunmore, Yale University Press, 2023. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=XvEdEAAAQBAJ. Acesso em: 29 jul. 2024.

CREMER, D. D.; KOOPMAN, J. Research: Using AI at Work Makes Us Lonelier and Less Healthy. Harvard Business Review, 2024. Disponível em: https://hbr.org/2024/06/research-using-ai-at-work-makes-us-lonelier-and-less-healthy. Acesso em: 29 jul. 2024.

DARCY, A. et al. Woebot: a chatbot for mental health. Journal of Medical Internet Research, 2019.

DURDEN, E.; PIRNER, M. C.; RAPOPORT, S. J.; WILLIAMS, A.; ROBINSON, A.; FORMAN HOFFMAN, V. Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent "Woebot." Journal of Medical Internet Research, v. 33, p. 100637–7, 2023.

ENHOLM, I. M. et al. Artificial Intelligence and Business Value: a literature review. Information Systems Frontiers, v. 24, n. 5, p. 1709-1734, 25 ago. 2021. Disponível em: https://link.springer.com/article/10.1007/s10796-021-10186-w. Acesso em: 29 jul. 2024.

EUROPEAN COMMISSION. Proposal for a Regulation on a European Approach for Artificial Intelligence. 2021. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

FARID, H. Deepfakes and the New Disinformation War. Scientific American, 2019.

GRAGLIA, M. A. V.; HUELSEN, P.; MUNIZ, A.; VALENTE, C. (org.). Inteligência Artificial: entenda como a IA pode impactar no mercado de trabalho e na sociedade. Rio de Janeiro: Brasport, 2024.

GRAGLIA, M. A. V.; LAZZARESCHI, N. A Indústria 4.0 e o Futuro do Trabalho: tensões e perspectivas. Revista Brasileira de Sociologia - RBS, v. 6, n. 14, 30 dez. 2018.

GRAGLIA, M. A. V.; LAZZARESCHI, N.; HUELSEN, P. G. V. As formas de resistência dos trabalhadores no contexto da indústria 4.0 e da inteligência artificial. Argumentum, Vitória, v. 13, n. 2, p. 193-207, maio/ago. 2021. Disponível em: https://doi.org/10.47456/argumentum.v13i2.27875.

GREENFIELD, B. AI chatbot prompted a 14-year-old's suicide, mom's lawsuit alleges: "We are behind the eight ball." Here's how to keep kids safe from new tech. Nova Iorque, Fortune, 2024. Disponível em: https://fortune.com/well/article/how-parents-can-keep-kids-safe-ai-chatbot-suicide-14-year-old-lawsuit/. Acesso em: 1 dez. 2024.

GROH, M. et al. Human detection of political speech deepfakes across transcripts, audio, and video. Nature Communications, v. 15, n. 1, p. 1–16, 2 set. 2024.

HETZNER, C. Gen Z men could ditch real women for AI, warns Ex-Google CEO Eric Schmidt after a tragic suicide involving a chatbot. Nova Iorque, Fortune, 2024. Disponível em: https://fortune.com/2024/11/28/gen-z-men-lonlineness-ai-google-eric-schmidt-suicide-sewell-setzer-sex-robots/. Acesso em: 1 dez. 2024.

HU, K. ChatGPT sets record for fastest-growing user base – analyst note. Nova Iorque, Reuters, 2023. Disponível em: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/. Acesso em: 29 jul. 2024.

LAGRANDEUR, K. The consequences of AI hype. AI and Ethics, 2023.

LAFFIER, J.; REHMAN, A. Deepfakes and Harm to Women. Journal of Digital Life and Learning, v. 3, n. 1, p. 1–21, 2023.

LAZZARESCHI, N; GRAGLIA, M. A. V. (Org.). Transformações no mundo do trabalho tensões e perspectivas. São Paulo: EDUC, 2023.

LIPPI, F. L.; ABILIO, C. C. C.; LIPPI, J. R.; GRAGLIA, M. A. V. Inteligência Artificial e saúde mental no Brasil: uma revisão sistemática da literatura. Contribuciones a las Ciencias Sociales, v. 17, n. 6, e7935-5, 2024. Disponível em: https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/7935. Acesso em: 29 jul. 2024.

MAHESHWARI, R. Top AI statistics and trends in 2023. Forbes, 2024. Disponível em: https://www.forbes.com/advisor/in/business/ai-statistics/. Acesso em: 29 jul. 2024.

MCKINSEY & COMPANY. Tech for Good: using technology to smooth disruption and improve well-being. 2019. Disponível em: https://www.mckinsey.com/featured-insights/future-of-work/tech-for-good-using-technology-to-smooth-disruption-and-improve-well-being. Acesso em: 29 jul. 2024.

MELLO, B. Adoção de IA cresceu 250% em 2023: o que há pela frente? Mundo do Marketing, 2024. Disponível em: https://www.mundodomarketing.com.br/adocao-de-ia-cresceu-250-em-2023-o-que-ha-pela-frente/. Acesso em: 29 jul. 2024.

MINH, D. et al. Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review, v. 55, p. 3503–3568, 2021.

MUSTAFA, A. N. The future of mathematics education: Adaptive learning technologies and artificial intelligence. International Journal of Science and Research Archive, v. 12, n. 1, p. 2594–2599, 2024.

NAVEED, H. et al. A Comprehensive Overview of Large Language Models. arXivLabs, 2023 [atualizado em 2024]. Disponível em: https://arxiv.org/abs/2307.06435. Acesso em: 29 jul. 2024.

NAZANIN, A. et al. What should we do with Emotion AI? Towards an Agenda for the Next 30 Years. In: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24). Nova Iorque, 2024. Disponível em: https://doi.org/10.1145/3678884.3689135. Acesso em: 1 dez. 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. The OECD Truth Quest Survey: Methodology and findings. OECD Digital Economy Papers, n. 369, Paris, 2024. Disponível em: https://doi.org/10.1787/92a94c0f-en. Acesso em: 29 jul. 2024.

PICARD, R. W. Affective Computing. MIT Press, 1997.

REUTERS FACT CHECK. Video does not show Joe Biden making transphobic remarks. 2023. Disponível em: https://www.reuters.com/article/fact-check/video-does-not-show-joe-biden-making-transphobic-remarks-idUSL1N34Q1IW/. Acesso em: 29 jul. 2024.

ROSA, J. Branding e inteligência artificial na publicidade de marca: estudo de caso do comercial “Gerações - 70 anos Volkswagen”. Trabalho de Conclusão de Curso (Graduação em Publicidade e Propaganda) – Universidade Federal do Rio de Janeiro. Rio de Janeiro, p. 83. 2024. Disponível em: https://pantheon.ufrj.br/handle/11422/24328. Acesso em: 29 jul. 2024.

SAMER, H.; AL-KHAZRAJI, S.; SALEH, H.; KHALID, A.; MISHKHAL, I. Impact of Deepfake Technology on Social Media: detection, misinformation and societal implications. Technology and Science, v. 23, 2023. Disponível em: http://www.epstem.net/tr/download/article-file/3456697. Acesso em: 29 jul. 2024.

SECURITYHERO.IO. 2023 State of Deepfakes: realities, threats, and impact. 2023. Disponível em: https://www.securityhero.io/state-of-deepfakes/. Acesso em: 29 jul. 2024.

SENADO FEDERAL. Relator apresenta relatório atualizado sobre regulamentação da IA. Brasília, 2024. Disponível em: https://www12.senado.leg.br/noticias/materias/2024/07/04/relator-apresenta-relatorio-atualizado-sobre-regulamentacao-da-ia. Acesso em: 29 jul. 2024.

SINGH, O. Artificial intelligence in the era of ChatGPT – opportunities and challenges in mental health care. Indian Journal of Psychiatry, v. 65, n. 3, p. 297, 2023.

TANG, P. M.; KOOPMAN, J.; KE, M. M.; DE CREMER, D.; ZHANG, J. H.; REYNDERS, P. et al. No person is an island: unpacking the work and after-work consequences of interacting with artificial intelligence. Journal of Applied Psychology, v. 108, n. 11, p. 1766–89, 2023.

WANG, Y. et al. A systematic review on affective computing: emotion models, databases, and recent advances. Information Fusion, v. 83-84, p. 19–52, 2022.

WORLD HEALTH ORGANIZATION. Ethics and governance of artificial intelligence for health. WHO, 2021. Disponível em: https://www.who.int/publications/i/item/9789240029200. Acesso em: 29 jul. 2024.