A ANTROPOMORFIZAÇÃO, O ANTROPOMORFISMO E A EMPATIA ARTIFICIAL COMO MODULADORES DA ACEITAÇÃO E RISCOS NA INTERAÇÃO HUMANO-MÁQUINA

José Carlos Rodrigues[1]

THE ANTHROPOMORPHIZATION, THE ANTHROPOMORPHISM, AND ARTIFICIAL EMPATHY AS MODULATORS OF ACCEPTANCE AND RISKS IN HUMAN-MACHINE INTERACTION

LA ANTROPOMORFIZACIÓN, EL ANTROPOMORFISMO Y LA EMPATÍA ARTIFICIAL COMO MODULADORES DE LA ACEPTACIÓN Y LOS RIESGOS EN LA INTERACCIÓN HUMANO-MÁQUINA

RESUMO

A evolução tecnológica resultou em uma interação cada vez mais frequente entre humanos e máquinas. Neste contexto, objetivando experiências mais fluidas e empáticas com seus utilizadores humanos, máquinas passaram incorporar características em sua forma e padrão de interação que remetem a atributos antropomórficos e/ou antropomorfizados (ou, zoomórficos ou zoomorfizados), impactando a interação e resultando em percepções de intencionalidade, empatia, capacidade de compreender e expressar emoções por parte dos seres humanos que com elas interagem. Este artigo teórico examina como a antropomorfização, o antropomorfismo e a empatia artificial modulam a aceitação e os riscos na interação humano-máquina. Discute-se como a simulação de emoções por robôs e inteligências artificiais pode facilitar a aceitação destas tecnologias, mas também levanta questões éticas sobre autenticidade e manipulação emocional. Embora a emulação de emoções possa criar vínculos mais fortes entre humanos e máquinas, as emoções expressas pelos seres artificiais são programadas, não genuínas, o que pode gerar expectativas irreais e potencial manipulação do comportamento humano, preocupação particularmente relevante em contextos vulneráveis como saúde mental e educação.

Palavras-chave: Inteligência Artificial; Interação humano-máquina; empatia artificial; antropomorfização; antropomorfismo.

ABSTRACT

Technological evolution has resulted in increasingly frequent interactions between humans and machines. In this context, aiming for more fluid and empathetic experiences with their human users, machines have begun to incorporate characteristics in their form and interaction patterns that resemble anthropomorphic and/or anthropomorphized attributes (or zoomorphic or zoomorphized ones), impacting the interaction and resulting in perceptions of intentionality, empathy, and the ability to understand and express emotions by the humans who interact with

[1] Doutorando em Comunicação e Práticas de Consumo (PPGCOM-ESPM), Mestre em Comportamento do Consumidor (MPCC-ESPM), especialista em storytelling, neurociências e impacto da tecnologia no comportamento humano. Email: acad@jcrodrigues.com.br

them. This theoretical article examines how anthropomorphization, anthropomorphism, and artificial empathy modulate acceptance and risks in human-machine interaction. It discusses how the simulation of emotions by robots and artificial intelligences can facilitate the acceptance of these technologies, but also raises ethical questions about authenticity and emotional manipulation. Although the emulation of emotions can create stronger bonds between humans and machines, the emotions expressed by artificial beings are programmed, not genuine, which can generate unrealistic expectations and potential manipulation of human behavior, a concern particularly relevant in vulnerable contexts such as mental health and education.

Keywords: Artificial Intelligence; Human-machine interaction; Artificial empathy; Anthropomorphization; Anthropomorphism.

RESUMEN

La evolución tecnológica ha resultado en una interacción cada vez más frecuente entre humanos y máquinas. En este contexto, con el objetivo de lograr experiencias más fluidas y empáticas con sus usuarios humanos, las máquinas han comenzado a incorporar características en su forma y patrones de interacción que remiten a atributos antropomórficos y/o antropomorfizados (o zoomórficos o zoomorfizados), impactando la interacción y generando percepciones de intencionalidad, empatía y capacidad de comprender y expresar emociones por parte de los seres humanos que interactúan con ellas. Este artículo teórico examina cómo la antropomorfización, el antropomorfismo y la empatía artificial modulan la aceptación y los riesgos en la interacción humano-máquina. Se discute cómo la simulación de emociones por robots e inteligencias artificiales puede facilitar la aceptación de estas tecnologías, pero también plantea cuestiones éticas sobre autenticidad y manipulación emocional. Aunque la emulación de emociones puede crear vínculos más fuertes entre humanos y máquinas, las emociones expresadas por los seres artificiales son programadas, no genuinas, lo que puede generar expectativas irreales y potencial manipulación del comportamiento humano, una preocupación particularmente relevante en contextos vulnerables como la salud mental y la educación.

Palabras clave: Inteligencia Artificial; Interacción humano-máquina; Empatía artificial; Antropomorfización; Antropomorfismo.

Introdução

A evolução da humanidade tem sido marcada por uma contínua interação (e coevolução) com a ciência e tecnologia (Rip, 2002), resultando em reconfigurações de como os seres humanos percebem e modificam elementos de seu ambiente, incluindo a concepção (ou dicotomia) de *sujeitos* e *objetos* discutida no projeto de modernidade (Latour, 1994), em que a noção de que *sujeitos*, representados pelos humanos, seriam agentes de conhecimento, cultura

e significado, enquanto, objetos, entidades não-humanas, vistas como passivas, naturais e sem agência, como animais e máquinas.

Além de uma visão funcionalista da integração e interação física e cognitiva dos seres humanos em simbioses cibernéticas, um outro aspecto dessa transformação envolve a incorporação de habilidades emocionais em máquinas e robôs na busca por maior aceitação e adoção por parte dos humanos e no estabelecimento de vínculos empáticos para com os tais *objetos* (Picard, 2000). Na realidade, a própria compreensão sobre o cognitivo considera o *pensar* e o *sentir* como intrinsecamente conectados – "se nós desejamos projetar um dispositivo que 'pense' no sentido de emular o cérebro humano, então ele também não deveria 'sentir'" (Picard, 2000, p.12).

A capacidade de expressar e interpretar emoções humanas não apenas redefine a funcionalidade das máquinas, mas também levanta questões sobre a natureza da empatia e da conexão emocional em um mundo em que "A relação entre animal e máquina é colocada em novos termos pela cibernética", com "as máquinas funcionam de modo semelhante aos organismos biológicos e [com] esse funcionamento [baseando-se] na troca de mensagens com o ambiente com o objetivo de diminuir a entropia" (Regis, 2023, p. 152).

Por outro lado, a inserção da capacidade de máquinas emularem emoções também provoca debates éticos necessários. A possibilidade de que robôs possam manipular emoções humanas ou criar laços afetivos que são essencialmente artificiais levanta questões sobre a autenticidade das interações e os possíveis impactos psicológicos nos interlocutores humanos. Wiener (1966) já alertava sobre os perigos de uma sociedade excessivamente dependente de máquinas, sugerindo que a linha entre o uso benéfico e o abuso da tecnologia pode ser tênue.

O presente artigo busca não apenas analisar as possibilidades tecnológicas e os benefícios da incorporação de emoções em robôs e Inteligências Artificiais (IA), mas também discutir as implicações éticas e sociais dessa prática, buscando fomentar a discussão sobre os desafios e oportunidades que surgem com a introdução de *robôs emocionalmente inteligentes* (Lee, 2019), ou, capazes de interpretar interações e expressar sinais antropomorfizados que emulem emoções percebidos como autênticos.

Ao examinar como a capacidade de expressar emoções emuladas pode moldar a empatia humana e a interação social, o artigo contribui para o debate sobre o futuro das relações entre humanos e máquinas em um mundo cada vez mais *tecnificado*.

Leitura Flutuante - Revista do Centro de Estudos em Semiótica e Psicanálise (São Paulo, *online*).
ISBN: 2175-7291. V.17 - Nº 1 2025 São Paulo - Brasil

Os tipos de IA e a incorporação de emoções emuladas em seres artificiais

A evolução das IAs tem avançado em duas abordagens, sob a perspectiva de aprendizagem de máquina: a IA discriminativa e a IA generativa (Jebara, 2012).

A IA discriminativa é projetada para classificar dados e tomar decisões com base nas informações de entrada a partir de modelos discriminativos que aprendem a identificar e distinguir entre diferentes classes de dados (Goodfellow, Bengio & Courville, 2016). Por sua vez, modelos generativos são projetados para forjar a distribuição conjunta de dados e têm a capacidade de gerar novas amostras que não estavam presentes no conjunto original de dados de treinamento (Feuerriegel, Hartmann, Janiesch & Zschech, 2019).

IAs generativas, em particular, vem obtendo considerável exposição na sociedade pela facilidade e naturalidade na interação com público em geral, e pela incorporação de elementos e comportamentos que emulam características humanas, como interjeições, pausas na fala e outras características que tornam as interações mais humanizadas, melhorando a experiência do usuário em contextos sociais e de marketing (Huang & Rust, 2018).

O constructo *empatia* é trazido para identificar a afetação das relações entre seres humanos e artificiais. Ainda que a emoção possa ser experimentada de forma individual, é na "capacidade de formar uma representação incorporada do estado emocional de outra pessoa, ao mesmo tempo estando ciente do mecanismo causal que induziu esse estado emocional" (Asada, 2015) que residiria o afeto (afetação) entre tais seres. Toma-se também, aqui, que a empatia não se refere apenas à percepção e entendimento dos estados mentais do outro, mas também à resposta a tal percepção (Barker, 2008, p.141).

Regis (2022) discute a importância da integração entre cognição e afeto na comunicação humano-máquina, argumentando que a simulação de emoções humanas pelas máquinas pode não apenas facilitar a interação, mas também influenciar a percepção dos humanos, contribuindo para uma nova forma de subjetividade nas relações com a tecnologia. Gabriel (2020), por sua vez, explora como a incorporação de emoções em robôs e assistentes virtuais pode aumentar a empatia dos usuários, tornando as interações mais eficazes.

A expressão de empatia de robôs na relação com humanos é apresentada como vital nos contextos em que tais seres artificiais prestarão serviços para os seres humanos (James,

Leitura Flutuante - Revista do Centro de Estudos em Semiótica e Psicanálise (São Paulo, *online*).
ISBN: 2175-7291. V.17 - Nº 1 2025 São Paulo - Brasil

Watson & Macdonald, 2018; Park & Whang, 2022), como nos setores de atendimento ao cliente, educação, cuidados de saúde e assistência pessoal, circunstâncias em que tais dispositivos são referenciados como *robôs sociais*, em tradução livre, definidos como:

> "[...] robôs que são especificamente projetados para interagir e se comunicar com pessoas, seja de forma semiautônoma ou autônoma (ou seja, com ou sem uma pessoa controlando o robô em tempo real), seguindo normas comportamentais que são típicas da interação humana". (Bartneck & Forlizzi, 2004)

Ao demonstrar comportamentos emocionais, esses robôs podem proporcionar conforto e apoio emocional a pessoas que podem estar isoladas ou vulneráveis. Estudos sugerem que a presença de robôs sociais emocionalmente expressivos pode contribuir para reduzir sentimentos de solidão e ansiedade, promovendo uma sensação de bem-estar entre os usuários (Breazeal, 2003; Gabriel, 2020).

Inspirados em mecanismos emocionais humanos – como, por exemplo, o *back-channel feedback*[2] (Christian, 2011, p.221) - esses modelos utilizam *EmotionPrompt* - técnica utilizada em modelos de inteligência artificial generativa que incorpora estímulos emocionais para melhorar a qualidade e a naturalidade das respostas geradas -e outras técnicas para ajustar a resposta da IA de forma a simular uma resposta emocional humana, o que pode aumentar a eficácia em tarefas colaborativas e na resolução de problemas (Zhu, Sudarshan, Kow & Ong 2024).

Quando a corporeidade é adicionada, a colaboração entre humanos e estes seres artificias se fortalece (Zhou & Tian, 2020), resultando, adicionalmente, em uma experiência emocional mais positiva para os humanos.

Válido destacar que não se busca discutir uma eventual efetiva incorporação das emoções em *coisas* ou *objetos* (robôs e IAs incorpóreas) como parte integral da transferência de capacidades cognitivas, às máquinas, na visão de Moravec (1988), mas em como a - relativamente mais simples - simulação de tais estados mentais e comportamentos responsivos

[2] O *back-channel feedback* é uma resposta do receptor da mensagem geralmente em um momento de comunicação unidirecional, podendo ser tanto verbal quanto não verbal (como balançar a cabeça concordando, por exemplo) com função fática, ou seja, criar uma conexão entre emissor e receptor, seja para iniciar, manter ou interromper a comunicação. (KNUDSEN et al., 2020)

afetam a interação com interlocutores humanos, afetando-os e ao processo comunicacional em si (Picard, 2000; Breazeal, 2003).

Antropomorfização e Antropomorfismo de seres artificiais

Em algum momento de sua carreira, todo psicólogo deve escrever uma versão da sentença "O ser humano é o único animal que....." (Christian, 2011, p.27). A compreensão do que completa esta sentença possui uma volatilidade que acompanha não somente a interpretação do que *é* o/um ser humano, mas de como ele foi capaz de emular suas funções em dispositivos antes criados à sua imagem e semelhança, como computadores e robôs.

Na popularização de sistemas computacionais que buscam replicar processos de tomada de decisão de seres humanos, vê-se um crescimento da adoção de ferramentas de IA e sua consequente integração em diferentes áreas do conhecimento e no cotidiano, ainda que as percepções quanto à tal tecnologia alternem-se entre visões positivas e negativas (Gerlich, 2023).

Discussões sobre a evolução da IA vão além de suas capacidades discriminativas ou generativas, mas também envolvem sua corporificação, em tradução livre, mecanicista e/ou fenomenológica (Sharkey & Ziemke, 2001), sendo respectivamente entendidas como a existência de um corpo/forma físico da IA (referenciado, por vezes, como *robôs*, mesmo que estes não necessariamente incorporem capacidades de IA) e a existência de uma percepção mental e adequações comportamentais.

Park e Whang (2022) discorrem sobre a modulação de características físicas (forma) e condutas (comportamento) como parte do processo da interação humano-robô (HRI) (Figura 1), tratando de forma independente aspectos motores (relacionados à forma) e cognitivos (relacionados à ideia de mente).

Leitura Flutuante - Revista do Centro de Estudos em Semiótica e Psicanálise (São Paulo, *online*).
ISBN: 2175-7291. V.17 - Nº 1 2025 São Paulo - Brasil

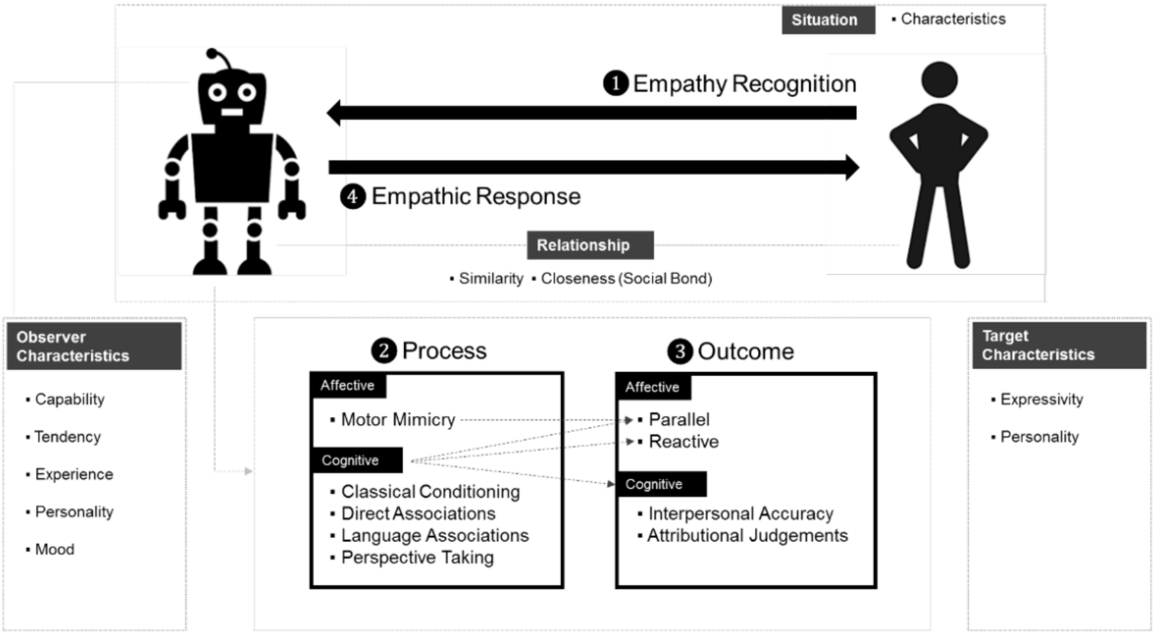

Figura 1 - Modelo conceitual da empatia na HRI
Fonte: Park e Whang (2022)

Considera-se que a corporificação da IA – em qualquer nível de antropomorfização, de mecatrônicos a autômatos - não seria uma premissa necessária, haja vista pode assumir diversas formas, não necessariamente materiais/físicas. Para entender sua essência, fundamentado no questionamento do princípio filosófico *res cogitans* e *res extensa* - separação entre mente e corpo - proposto por Descartes (1996), podemos tratar cada elemento como distinto e independente.

Isto quer dizer que não seria a existência de um *corpo* (ou partes dele) o determinante da criação de vínculo de agentes artificiais com humanos, sendo relevante, mas não suficiente ou necessário (Pelau, Dabija & Ene, 2021). Cerf (2003), por exemplo, trata o termo *robô* como "programas que executam funções, ingerem entradas e produzem saídas que têm um efeito perceptível" (tradução nossa), ainda que não haja um elemento ou efeito físico.

Assim, por conta desta independência da corporificação do agente artificial – referenciando robôs e IAs incorpóreas –, e potenciais interpretações divergentes entre *Robôs* e *Inteligência Artificial*, tais agentes serão doravante referenciados como *seres artificiais*, quando da percepção de agência e intencionalidade por parte dos interlocutores humanos.

Isto abre espaço para se discutir a consideração de dois aspectos destes seres artificiais: a antropomorfização e o antropomorfismo.

Leitura Flutuante - Revista do Centro de Estudos em Semiótica e Psicanálise (São Paulo, *online*).
ISBN: 2175-7291. V.17 - Nº 1 2025 São Paulo - Brasil

Alguns/mas autores/as mesclam aspectos físicos e mentais quando se referem ao termo *antropomorfização*, utilizando-o como uma representação holística de características humanas físicas e mentais (Scorici, Schultz & Seele, 2022; Röhe & Santaella, 2023). Neste texto, porém, considera-se uma diferenciação entre a representação física (forma) do ser artificial – *antropomorfização* – e a expressão (simulação) de estados mentais e capacidades empáticas deste ser – *antropomorfismo* -, independente de sua forma ou mesmo da existência de uma representação física determinada.

Toma-se o antropomorfismo como "a tendência a atribuir características humanas ou similar a humanas e agentes não humanos" (Epley, Waytz & Cacioppo, 2007), incluindo estados mentais (Lee, Lee & Sah, 2019), enquanto a *antropomorfização* como a representação de aparência similar ou próxima a de um ser humano por meio de um corpo ou partes que remetam a este (como presença de cabeça, face e/ou membros superiores/inferiores). Esta representação poderia se dar com dispositivos tangíveis (físicos) ou por meio de telas e outros aparatos de projeção de imagem (como óculos de realidade aumentada e virtual, por exemplo).

Embora em um contexto muito anterior ao desenvolvimento das tecnologias modernas, Descartes (1996) oferece uma base filosófica para compreender o desconforto que pode acompanhar o antropomorfismo de máquinas. Ao defender a distinção clara entre o corpo mecânico e a mente racional, argumenta que, embora possamos construir autômatos que imitem as funções corporais humanas, a razão, e por extensão as emoções genuínas, seriam exclusivas dos seres humanos. Esse ponto de vista cartesiano ainda ressoa nas discussões contemporâneas sobre os limites do que as máquinas podem realmente alcançar em termos de humanidade, sugerindo uma linha divisória entre imitação e essência, ou uma flexibilidade mental para encarar tais máquinas como simulacros humanos ou animais.

Não se considera, contudo, retomar aqui a visão dualística entre *corpo* e *não-corpo*, mas de um *continuum* que considera o que se propõe como *nível de vitalização* do (então) *objeto* a partir de sua forma física (reconhecimento, pelo/a interlocutor/a, de partes físicas que simulariam cabeças e membros) e/ou capacidades de expressar-se por meio de outras interfaces. A vitalização expressa a percepção de vida, intencionalidade e empatia (como resposta a estímulos) de um ser por outros, reunindo aspectos da antropomorfização e do antropomorfismo ou, mais além, do animismo, crença, segundo a qual, um grande número de entidades não humanas possui uma alma (Castro, 2019).

Autenticidade e Manipulação nas Interações Humano-Máquina

Apesar dos potenciais benefícios no fortalecimento empático da relação humano-máquina e na maior aceitação de robô sociais com a emulação de emoções (James, Watson & Macdonald, 2018), há que se considerar potenciais riscos de confiar excessivamente em máquinas que imitam características humanas.

Embora não trate diretamente da incorporação de emoções, Wiener (1966) alerta que a dependência crescente de tais tecnologias pode levar a uma falsa sensação de empatia e confiança nas máquinas, com potenciais consequências negativas para a sociedade. Nem dizer sobre as preocupações técnicas e filosóficas trazidas por Latour (1994) na reflexão sobre a tradicional separação entre natureza e cultura e na criação de novas formas de subjetividade e interação.

Quando seres artificiais são projetados para imitar emoções humanas, pode-se questionar o quanto tais simulações seriam capazes de manipular a resposta emocional do usuário, levando os humanos a atribuir estados emocionais *reais* a máquinas, criando uma interação baseada em uma falsa percepção de empatia, o que pode gerar expectativas irreais e distorcer a natureza da relação, seguindo o princípio de *mediação radical* de Grusin (2015), como um evento ontológico que antecede a distinção entre sujeitos e objetos.

Quando as emoções humanas são replicadas por seres artificiais, há um risco de uma confiança indevida em sistemas que são, em última análise, programados para reagir de maneiras predeterminadas (Gabriel, 2020). Essa situação compromete a percepção da autenticidade nas relações formadas com essas máquinas, uma vez que as respostas emocionais dos robôs não são fruto de uma experiência vivida, mas de um algoritmo projetado para produzir empatia (Turkle, 2011).

A manipulação emocional por robôs é outro fator a se discutir. A capacidade dos seres artificiais de simular emoções pode ser usada para influenciar os usuários de maneiras que não seriam possíveis se as emoções simuladas fossem reconhecidas como artificiais (Bakir, Laffer, Mcstay, Miranda & Urquhart, 2024).

Isso pode ser particularmente perigoso em contextos vulneráveis, como na saúde mental ou em situações de suporte emocional, onde os usuários podem depender da empatia percebida dos robôs para conforto. A ilusão de empatia pode, assim, ser explorada para

manipular o comportamento humano, levantando sérias questões éticas sobre o uso dessas tecnologias.

E, embora Ienca (2023) proponha a *intencionalidade* como um dos quatro fatores para manipulação (juntamente à assimetria de resultados, não-transparência e violação da autonomia), as regras lógicas que levam um ser artificial a dar uma diretiva a um humano não necessariamente contemplariam um comportamento deliberativo, podendo ser resultado de conclusões não benéficas (ao humano), seja como parte de lógica algorítmica natural, potencialmente encapsulada nas caixas pretas codificadas (Pasquale, 2015, apud Kubler, 2015) ou da programação deliberada de seus desenvolvedores.

Percepção de intencionalidade e resposta humana na interação com seres artificiais

A percepção de intencionalidade em seres artificiais constitui um fator determinante na modulação das respostas humanas durante interações com tais entidades (Ienca, 2023). Quando humanos atribuem intenções, desejos e crenças a robôs ou sistemas de IA, tendem a engajar-se em comportamentos sociais mais complexos e emocionalmente investidos, mesmo sabendo racionalmente que estão interagindo com máquinas programadas (Waytz, Gray, Epley & Wegner, 2010).

Este fenômeno, tratado por Di Stasio e Miotti (2024) como *teoria da mente artificial*, leva os usuários a interpretar ações algorítmicas como decisões deliberadas, atribuindo responsabilidade moral e capacidade de julgamento a sistemas que operam por meio de processos computacionais predeterminados.

Estudos experimentais demonstram que a mera sugestão de que um sistema possui autonomia decisória aumenta significativamente a confiança depositada pelos usuários, mesmo quando o desempenho objetivo permanece inalterado (Złotowski, Yogeeswaran & Bartneck, 2017).

A intensidade com que humanos respondem à percepção de intencionalidade varia conforme o contexto da interação e características individuais do usuário. Em ambientes terapêuticos ou educacionais, por exemplo, a atribuição de intencionalidade a assistentes

Leitura Flutuante - Revista do Centro de Estudos em Semiótica e Psicanálise (São Paulo, *online*).
ISBN: 2175-7291. V.17 - Nº 1 2025 São Paulo - Brasil

artificiais pode potencializar resultados positivos através do efeito placebo social, onde a crença na capacidade empática do sistema catalisa mudanças comportamentais reais no usuário (Nass & Moon, 2000). Paradoxalmente, esta mesma percepção pode gerar expectativas irrealistas sobre as capacidades do sistema, resultando em frustrações quando limitações técnicas se manifestam. Indivíduos com maior tendência à antropomorfização demonstram respostas emocionais mais intensas a falhas de sistemas percebidos como intencionais, experimentando sentimentos de traição ou abandono que não seriam despertados por ferramentas consideradas meramente instrumentais (Damiano & Dumouchel, 2018).

As implicações éticas desta dinâmica são particularmente relevantes quando consideramos populações vulneráveis ou situações de dependência emocional. A percepção de intencionalidade pode criar assimetrias de poder onde o usuário desenvolve uma relação de confiança unilateral com um sistema incapaz de reciprocidade genuína, mas programado para simular tal reciprocidade (Coeckelbergh, 2019). Este descompasso entre a realidade técnica e a experiência fenomenológica do usuário levanta questões sobre consentimento informado e manipulação emocional, especialmente quando designers exploram deliberadamente vieses cognitivos humanos para intensificar a percepção de agência em sistemas artificiais. A transparência sobre a natureza algorítmica das respostas e a educação dos usuários sobre os mecanismos subjacentes às interações tornam-se, portanto, imperativos éticos fundamentais no desenvolvimento responsável de tecnologias que simulam intencionalidade (Turkle, 2016).

Considerações finais

Este artigo ressalta a importância de se compreender as implicações da incorporação de emoções em seres artificiais, especialmente no contexto de suas interações com humanos (robôs sociais), independente da corporificação destes seres. Tanto robôs físicos quanto assistentes virtuais podem ser programados para simular emoções, facilitando a formação de vínculos com os usuários.

Embora a simulação de emoções por robôs possa facilitar a formação de laços entre humanos e máquinas, aumentando a aceitação dessas tecnologias, também levanta questões sobre a autenticidade dessas relações, pois poderia haver percepções, por parte dos humanos, sobre intencionalidade e consciência do ser artificial, o que poderia modular seu próprio comportamento e julgamento.

Vê-se vestígios desta interação em estudos que avaliam a dinâmica entre seres humanos e Inteligências Artificiais Generativas, como em Nawar (2024) e Lowe (2025). Considera-se até que a educação (ou *polidez*, em tradução livre do termo original em inglês, *politeness*) para com tais seres artificiais exerceria um efeito reverso, fomentando tal comportamento dos seres humanos entre si.

Há que se refletir sobre a construção desses laços, já que as emoções expressas pelos seres artificiais são, em última análise, programadas, e não fruto de uma experiência emocional genuína e vitalizada, levantando questões sobre a potencial manipulação emocional e a confusão entre o real e o artificial. A ilusão de empatia, seja por um robô físico ou um assistente virtual não-antropomórfico (ambos referidos como *seres artificias* por conta da sensação de presença[3] nas interações), poderia ser usada para influenciar o comportamento humano de maneiras que não seriam possíveis se os tais humanos reconhecessem essas expressões como artificiais.

Tal possibilidade é particularmente relevante com utilizadores humanos e/ou em contextos em que se desenvolva uma relação de confiança entre estes utilizadores e os seres artificiais, como na educação, saúde mental e suporte emocional.

As manipulações emocionais realizadas por seres artificiais podem ocorrer tanto de forma pré-programada, em que seus desenvolvedores inserem intencionalmente respostas emocionais específicas destinadas a influenciar o comportamento do usuário, como podem surgir naturalmente a partir dos algoritmos, à medida que estes aprendem e adaptam suas respostas com base em grandes volumes de dados e interações anteriores.

Ambas as formas apresentam desafios éticos significativos, pois, independentemente da origem da manipulação, o usuário pode ser influenciado sem estar plenamente consciente de que está interagindo com uma simulação emocional artificial.

À medida que as máquinas se tornam mais capazes de interpretar e responder a sinais emocionais, faz-se necessária uma avaliação das normas sociais e éticas que governam tais interações, de modo a assegurar que o desenvolvimento dessas tecnologias contribua positivamente para a sociedade, sem comprometer a integridade física ou mental dos *seres baseados em carbono* que interagem com os *seres baseados em silício*.

[3] Com base na Teoria da Presença Social (Short, Williams & Christie, 1978), considera-se como "a sensação de estar com outros [indivíduos] em uma ação coordenada" (tradução nossa)

REFERÊNCIAS

ASADA, Minoru. Towards artificial empathy: how can artificial empathy follow the developmental pathway of natural empathy?. International Journal of Social Robotics, v. 7, p. 19-33, 2015. Disponível em: https://link.springer.com/content/pdf/10.1007/s12369-014-0253-z.pdf. Acesso em 15-jan-2025

BAKIR, Vian; LAFFER, Alexander; MCSTAY, Andrew; MIRANDA, Diane; URQUHART, Lachlan. On Manipulation by Emotional AI: UK Adults' Views and Governance Implications. Frontiers in Sociology, v. 9, p. 1339834, 2024. Disponível em: https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2024.1339834/full. Acesso em 20-jan-2025

BARKER, Robert Lee. *The Social Work Dictionary*, Washington, DC: NASW Press. 2008

BREAZEAL, Cynthia. Toward sociable robots. Robotics and autonomous systems, v. 42, n. 3-4, p. 167-175, 2003. Disponível em: https://www.sciencedirect.com/science/article/pii/S0921889002003731?casa_token=dLBCIKEVsyIAAAAA:6RN4DQ9TAADG3RBxBR0EE4pjVeoMUegpkQMIpyC1DvScvwuWXQ11p1DUi2uxFDPh2S6qtWTmqqg. Acesso em 20-dez-2024

CASTRO, Teresa. O regresso do animismo. Cinema, efeitos de presença e imagens que agem. Pós-Fotografia, Pós-Cinema, Novas configurações das imagens., 2019. Disponível em: https://hal.science/hal-02410568/document. Acesso em 10-dez-2024

CERF, Vinton G. What's a robot?. Communications of the ACM, v. 56, n. 1, p. 7-7, 2013. Disponível em: https://dl.acm.org/doi/pdf/10.1145/2398356.2398358. Acesso em 09-set-2025

CHRISTIAN, Brian. *O humano mais humano*: o que a inteligência artificial nos ensina sobre a vida. São Paulo: Editora Schwarcz, 2011.

COECKELBERGH, Mark. Artificial intelligence: Some ethical issues and regulatory challenges. Technology and regulation, v. 2019, p. 31-34, 2019. Disponível em: https://techreg.org/article/view/10999. Acesso em 20-dez-2024

DAMIANO, Luisa; DUMOUCHEL, Paul. Anthropomorphism in human–robot co-evolution. Frontiers in psychology, v. 9, p. 468, 2018. Disponível em: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00468/full. Acesso em 10-jan-2025

DESCARTES, René. *Discurso do Método*. São Paulo: Nova Cultural em 1996.

DI STASIO, Margherita; MIOTTI, Beatrice. Intelligent Agents at School—Child–Robot Interactions as an Educational Path. Education Sciences, v. 14, n. 7, 2024. Disponível em: https://www.mdpi.com/2227-7102/14/7/774. Acesso em 14-jan-2025

EPLEY, Nicholas; WAYTZ, Adam; CACIOPPO, John T. On seeing human: a three-factor theory of anthropomorphism. Psychological review, v. 114, n. 4, p. 864, 2007. Disponível em: https://pubmed.ncbi.nlm.nih.gov/17907867/. Acesso em 02-dez-2024

FEUERRIEGEL, S.; HARTMANN, J.; JANIESCH, C.; ZSCHECH, P. Generative AI. Business & Information Systems Engineering, v. 66, n. 1, p. 111-126, 2024. Disponível em: https://link.springer.com/article/10.1007/s12599-023-00834-7. Acesso em 04-dez-2024.

GABRIEL, Martha. *Inteligência artificial*: do zero ao metaverso. São Paulo: Editora Atlas, 2020. Edição Kindle.

GERLICH, Michael. Perceptions and acceptance of artificial intelligence: A multi-dimensional study. Social Sciences, v. 12, n. 9, p. 502, 2023. Disponível em: https://www.mdpi.com/2076-0760/12/9/502. Acesso em 17-dez-2024.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge, MA: MIT Press, 2016.

GRUSIN, Richard. Premediation: affect and mediality after 9/11. New York: Palgrave MacMillan, 2010. Disponível em: https://link.springer.com/book/10.1057/9780230275270. Acesso em 08-ago-2024.

HUANG, Ming-Hui; RUST, Roland T. Artificial intelligence in service. Journal of service research, v. 21, n. 2, p. 155-172, 2018. Disponível em: https://journals.sagepub.com/doi/10.1177/1094670517752459. Acesso em: 23-out-2024.

IENCA, Marcello. On artificial intelligence and manipulation. Topoi, v. 42, n. 3, p. 833-842, 2023. Disponível em: https://link.springer.com/article/10.1007/s11245-023-09940-3. Acesso em: 14-ago-2024

JAMES, Jesin; WATSON, Catherine Inez; MACDONALD, Bruce. Artificial empathy in social robots: an analysis of emotions in speech. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, 2018. p. 632-637. Disponível em: https://ieeexplore.ieee.org/document/8525652. Acesso em 20-ago-2024

JEBARA, Tony. *Machine learning*: discriminative and generative. New York: Springer Science & Business Media, 2012.

KUBLER, Kyle. *The black box society*: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

LATOUR, Bruno. *Jamais fomos modernos*. Rio de Janeiro: Ed. 34, 1994.

LEE, Kai-Fu. *Inteligência artificial*. Rio de Janeiro: Editora Globo S.A., 2019.

LEE, Sangwon; LEE, Naeun; SAH, Young. Perceiving a Mind in a Chatbot: Effect of Mind Perception and Social Cues on Co-presence, Closeness, and Intention to Use. International Journal of Human–Computer Interaction, v. 35, n. 11, p. 903-911, 2019. Disponível em: https://www.tandfonline.com/doi/full/10.1080/10447318.2019.1699748. Acesso em 29-jan-2025

LOWE, Daniel. Please and thank you, ChatGPT. The Gippsland Anglican, p. 20-21, 2025. Disponível em: https://search.informit.org/doi/abs/10.3316/informit.T202502060001230112384965 8. Acesso em: 24-fev-2025

MORAVEC, Hans. *Mind Children*: The Future of Robot and Human Intelligence. Cambridge: Harvard University Press, 1988.

NASS, Clifford; MOON, Youngme. Machines and mindlessness: Social responses to computers. Journal of social issues, v. 56, n. 1, p. 81-103, 2000. Disponível em: https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/0022-4537.00153. Acesso em 10-10-2024

NAWAR, Essraa. Do You Say Please or Thank You to ChatGPT? The Subtle Influence of Prompt Engineering on Digital Civility. 2024. Disponível em: https://digitalcommons.chapman.edu/librarian_articles/44/. Acesso em 10-jul-2024

PARK, Sung; WHANG, Mincheol. Empathy in human–robot interaction: designing for social robots. International Journal of Environmental Research and Public Health, v. 19, n. 3, p. 1889, 2022. Disponível em: https://pubmed.ncbi.nlm.nih.gov/35162909/. Acesso em 14-ago-2023.

PELAU, Corina; DABIJA, Dan-Cristian; ENE, Irina. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Computers in Human Behavior, v. 122, p. 106855, 2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S0747563221001783. Acesso em 14-nov-2024.

PICARD, Rosalind W. *Affective computing*. Cambridge: MIT press, 2000.

REGIS, Fátima. *Nós, ciborgues*: tecnologias de comunicação e subjetividade humano-máquina. 2. ed., revista, atualizada e ampliada. Curitiba: PUCPRESS, 2023.

RIP, Arie. Co-evolution of science, technology and society. Revisión experta del Bundesministerium Bildung, 2002. Disponível em: http://www.sciencepolicystudies.de/dok/expertise-rip.pdf. Acesso em: 30-ago-2024.

RÖHE, Anderson; SANTAELLA, Lucia. Confusões e dilemas da antropomorfização das inteligências artificiais. TECCOGS: Revista Digital de Tecnologias Cognitivas, n. 28, p. 67-75, 2023. Disponível em: https://revistas.pucsp.br/index.php/teccogs/article/view/67070/45078. Acesso em: 12-out-2024

SCORICI, Manuela; SCHULTZ, Mario D.; SEELE, Peter. Anthropomorphization and beyond. AI & society, 2022. Disponível em: https://link.springer.com/article/10.1007/s00146-022-01492-1. Acesso em 06-dez-2024.

SHARKEY, Noel E.; ZIEMKE, Tom. Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI?. Cognitive Systems Research, v. 2, n. 4, p. 251-262, 2001. Disponível em: https://www.sciencedirect.com/science/article/pii/S1389041701000365. Acesso em 10-jan-2025.

SHORT, John; WILLIAMS, Ederyn; CHRISTIE, Bruce. The social psychology of telecommunications. Contemporary Sociology. Vol. 7, no. 1. 1978. Disponível em: https://www.jstor.org/stable/2065899. Acesso em 25-ago-2023.

TURKLE, Sherry. *Alone Together*: Why We Expect More from Technology and Less from Each Other. New York: Basic Books, 2011.

TURKLE, Sherry. *Reclaiming conversation*: The power of talk in a digital age. Westminster: Penguin, 2016.

WAYTZ, A.; GRAY, K.; EPLEY, N.; WEGNER, D. M. Causes and consequences of mind perception. Trends in Cognitive Sciences, v. 14, n. 8, p. 383-388, 2010. Disponível em: https://dtg.sites.fas.harvard.edu/DANWEGNER/pub/Waytz%20et%20al%202010.pdf. Acesso em: 14-dez-2024.

WIENER, Norbert. *Cibernética e sociedade*: o uso humano de seres humanos. 2. ed. São Paulo: Cultrix, 1966.

ZHOU, Shujie; TIAN, Leimin. Would you help a sad robot? Influence of robots' emotional expressions on human-multi-robot collaboration. In: 2020 29th IEEE international conference on robot and human interactive communication (RO-MAN). IEEE, 2020. p. 1243-1250. Disponível em: https://ieeexplore.ieee.org/document/9223524. Acesso em: 13-dez-2024.

ZHU, Gaoxia; SUDARSHAN, Vidya; KOW, Jason; ONG; Ywe. Human-Generative AI Collaborative Problem Solving Who Leads and How Students Perceive the Interactions. IEEE Conference on Artificial Intelligence (CAI) 2024. Disponível em: https://arxiv.org/abs/2405.13048. Acesso em 11-jan-2025.

ZŁOTOWSKI, Jakub; YOGEESWARAN, Kumar; BARTNECK, Christoph. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. International Journal of Human-Computer Studies, v. 100, p. 48-54, 2017. Disponível em: https://www.sciencedirect.com/science/article/pii/S1071581916301768. Acesso em 15-jan-2025.