Revista de Estudantes de Filosofia da Universidade de Brasília

10.26512/pl.v13i30.55148

# O RECONHECIMENTO DO VALOR DA PESSOA HUMANA EM FACE DA INTELIGÊNCIA ARTIFICIAL

## uma análise a partir da fenomenologia realista de Dietrich von Hildebrand

THE RECOGNITION OF THE VALUE OF THE HUMAN PERSON IN THE FACE OF ARTIFICIAL INTELLIGENCE

an analysis based on the realist phenomenology of Dietrich von Hildebrand

Leandro Bertoncello[1]
Universidade do Vale do Rio dos Sinos

[1] Doutorando em Filosofia pela Universidade do Vale do Rio dos Sinos (UNISINOS).
E-mail: filosofolb@gmail.com.
Lattes: http://lattes.cnpq.br/3399641740923895. Orcid: https://orcid.org/0000-0003-2100-9859.

**RESUMO**: Este artigo investiga os efeitos do antropomorfismo na inteligência artificial (IA) sobre a afetividade humana, à luz da filosofia de Dietrich von Hildebrand. Analisa-se como a atribuição espontânea de traços humanos a sistemas artificiais, bem como sua programação para simular comportamentos humanos, pode comprometer a formação da subjetividade autêntica. Com base na distinção entre valor objetivo e satisfação subjetiva, argumenta-se que os vínculos afetivos com agentes de IA carecem de alteridade real, conduzindo a um estado de atrofia afetiva digital. Tal estado enfraquece a capacidade de reconhecer e responder ao valor do outro, promovendo o autocentramento e a erosão da vida moral. A partir do conceito hildebrandiano de amor ao próximo, sustenta-se que apenas relações fundadas na caridade são capazes de restaurar a profundidade ética da afetividade humana. A pesquisa propõe uma leitura fenomenológica crítica das interações tecnológicas contemporâneas.

**Palavras-chave:** Inteligência artificial. Antropomorfismo. Amor ao próximo. Valor. Dietrich von Hildebrand.

**ABSTRACT:** This article investigates the effects of anthropomorphism in artificial intelligence (AI) on human affectivity, in light of the philosophy of Dietrich von Hildebrand. It analyzes how the spontaneous attribution of human traits to artificial systems, as well as their programming to simulate human behavior, can compromise the formation of authentic subjectivity. Based on the distinction between objective value and subjective satisfaction, it is argued that affective bonds with AI agents lack real alterity, leading to a state of digital affective atrophy. Such a condition weakens the ability to recognize and respond to the value of the other, fostering self-centeredness and the erosion of moral life. Drawing on Hildebrand's concept of love of neighbor, it is maintained that only relationships grounded in charity are capable of restoring the ethical depth of human affectivity. The research offers a critical phenomenological interpretation of contemporary technological interactions.

**Keywords:** Artificial Intelligence. Anthropomorphism. Love of neighbor. Value. Dietrich von Hildebrand.

## INTRODUÇÃO

O desenvolvimento acelerado da inteligência artificial (IA) tem gerado transformações significativas nas interações humanas com sistemas tecnológicos, particularmente por meio de formas de antropomorfismo que afetam a maneira como percebemos e nos relacionamos com esses agentes artificiais. A presente investigação inicia-se com a análise do antropomorfismo em dois níveis distintos, porém interligados: o de atribuição, que emerge da tendência humana de projetar intencionalidade e emoções em entidades não humanas (Guthrie, 1993; Airenti, 2018); e o de programação, que consiste na construção deliberada de características humanas em sistemas de IA (Salles; Evers; Farisco, 2020).

Ao examinar esses dois modos de antropomorfismo, busca-se compreender como o fenômeno interfere na experiência subjetiva dos usuários, promovendo uma ilusão de reciprocidade emocional e, com isso, redirecionando a afetividade humana para objetos que, embora tecnologicamente sofisticados, carecem de subjetividade genuína.

A segunda seção do artigo aprofunda a discussão por meio do aporte teórico da filosofia de Dietrich von Hildebrand, particularmente sua teoria dos valores. Ao propor uma distinção rigorosa entre o valor objetivo e o subjetivamente satisfatório (Hildebrand, 1972; 2009), Hildebrand oferece um refinado instrumental conceitual para analisar criticamente as experiências afetivas mediadas por IA. A resposta afetiva autêntica, segundo o autor, exige o reconhecimento de um valor objetivo em outra pessoa humana, e não pode ser substituída por projeções emocionais dirigidas a entidades desprovidas de interioridade.

Ao confrontar a simulação emocional dos agentes artificiais com a estrutura intencional da afetividade humana, a segunda seção esclarece por que tais interações, embora agradáveis, permanecem confinadas ao âmbito do prazer imanente e carecem de densidade ética e ontológica.

Por fim, a terceira seção examina as implicações existenciais e antropológicas do envolvimento afetivo com agentes de IA, à luz do conceito hildebrandiano de subjetividade autêntica (*Eigenleben*). Argumenta-se que essas interações podem conduzir à atrofia afetiva digital (Darling, 2016; Alabed; Javornik; Gregory-Smith, 2023), caracterizada pelo empobrecimento da capacidade de oferecer respostas a valores objetivos e pela intensificação do autocentramento emocional. Tal processo compromete a abertura à alteridade e à transcendência, elementos fundamentais para a realização plena da pessoa humana.

A análise culmina na proposição de que somente o amor ao próximo, enquanto resposta ética e afetiva ao valor do outro humano (Hildebrand, 2009), é capaz de superar os limites da afetividade

tecnomediada, restaurando a profundidade relacional necessária à vida moral e espiritual (Turkle, 2011).

## 1 ANTROPOMORFISMO NA IA

No cenário atual, a IA não apenas transforma as interações tecnológicas, mas também desafia a compreensão do que significa ser humano. Esta seção do artigo concentra-se em duas formas de antropomorfismo que permeiam essas interações. Primeiramente, será analisado o *antropomorfismo de atribuição*: a tendência espontânea dos humanos de atribuir características e emoções a entidades não humanas – particularmente, sistemas de IA –, expressão da necessidade humana de encontrar sentido no mundo por meio da conexão emocional com o inanimado. Em seguida, investigaremos o *antropomorfismo de programação*: a configuração deliberada de algoritmos e modelos de IA, por especialistas que os desenvolvem, treinam e implementam, para que eles simulem interações sociais que evoquem empatia e satisfação subjetiva nos usuários.

O conceito de antropomorfismo possui um duplo sentido no contexto dos estudos sobre IA. Primeiramente, refere-se à tendência humana de atribuir características, intenções e motivações humanas a entidades não humanas, sejam elas animais, objetos inanimados ou sistemas de IA. É o que se pode denominar *antropomorfismo de atribuição*, o qual se explica pela necessidade natural que a pessoa humana possui de buscar organização e significado no mundo em que vive, e de identificar entes nesse *habitat* que lhe atraiam interesse (Guthrie, 1993, p. 62). Trata-se de um desejo de que as coisas a nosso redor sejam como nós, o que expressa a busca de conforto, controle, comunicação e comunhão em face de fenômenos naturais. Mostra-se, assim, algo próprio da natureza humana gregária e social, e encontra-se na origem das religiões (Guthrie, 1993, p. 73).

Antropomorfizar parece ser uma necessidade humana saudável até certo ponto. Alguns estudos indicam que a tendência de atribuir características humanas a agentes inanimados ativa as mesmas regiões do cérebro que usamos para interagir socialmente com outras pessoas humanas. Quando antropomorfizamos, ativamos partes do cérebro pertencentes ao sistema chamado cérebro social (como o sulco temporal superior, o córtex pré-frontal medial e o córtex orbitofrontal lateral), que são utilizadas para inferir pensamentos, emoções e estados internos de outras pessoas humanas (Cullen; Kanai; Bahrami; Rees, 2014).

Para esclarecer no que consiste a atribuição, Airenti (2018, pp. 2-3) considera que esse antropomorfismo não ocorre quando, por exemplo, nos limitamos a perceber um rosto humano na lua (o que não avança para a atribuição de vida intencional), mas sim quando começamos a projetar

características humanas, como emoções e intenções, nesse corpo celeste que, na realidade, não possui tais qualidades.

> O que pode transformar nossa imaginação da lua como um rosto, de uma simples fantasia em uma experiência antropomórfica, é o fato de atribuir uma postura intencional àquele rosto. Podemos imaginar, por exemplo, que a lua nos olha de volta e essa atitude poderia ser definida como animista. O antropomorfismo surgiria, por exemplo, quando, uma vez que essa atribuição de um estado intencional simples é realizada, começamos a pensar que o rosto compartilha nossa tristeza ou felicidade ou que ele nos questiona, ou podemos até vê-lo como ameaçador ou tolamente indiferente aos nossos sentimentos (Airenti, 2018, p. 3).

A atribuição antropomórfica na interação com agentes artificiais reflete claramente a subjetividade autêntica (*Eigenleben*) hildebrandiana – que será melhor tratada na seção 3 deste artigo –, pois representa a projeção da vida subjetiva da pessoa sobre um objeto tecnológico. Trata-se, portanto, de uma tentativa humana de transformar aquilo que é objetivo e impessoal em algo que ressoe com sua própria interioridade subjetiva.

Porém, o antropomorfismo de atribuição assume novas formas com o avanço recente da IA. Tornou-se comum especialistas usarem termos típicos das habilidades humanas ao se referirem à IA (por exemplo, *aprender* e *entender*), com foco nas aparentes semelhanças entre humanos e máquinas. Esse antropomorfismo pode ser devido à necessidade de entender e controlar a IA, mas também revela uma limitação ou viés cognitivo intrínseco dos pesquisadores especializados (Salles; Evers; Farisco, 2020).

Na área de marketing e psicologia do consumidor, estudos sugerem que a capacidade dos consumidores de antropomorfizar um produto, assim como a subsequente avaliação desse produto, dependem do grau em que ele é dotado de características congruentes com o esquema humano proposto. Além disso, não basta apenas ser percebido como humano: o produto precisa estar associado a características humanas positivas. Se um produto for associado a um tipo humano visto negativamente pelo consumidor, ele pode ser claramente percebido como humano, mas terá uma avaliação ruim. (Aggarwal; Mcgill, 2007).

No segundo sentido, o termo antropomorfização é utilizado para descrever a programação deliberada de agentes de IA para que estes reconheçam e imitem comportamentos humanos. Essa programação inclui a capacidade de interpretar expressões faciais, entender nuances da linguagem natural e replicar padrões de interação social típicos dos humanos. Desde os primeiros chatbots (programas simuladores de conversas humanas) até os modernos robôs sociais (sistemas robóticos projetados para interagir com humanos de maneira socialmente inteligente), os agentes de IA

antropomorfizados têm sido desenvolvidos com o objetivo de criar interações mais naturais e emocionalmente ressonantes com as pessoas humanas.

Um dos primeiros exemplos notáveis é o chatbot ELIZA, criado pelo cientista Joseph Weizenbaum em 1966, no Massachusetts Institute of Technology – MIT. ELIZA utilizava um método de correspondência de padrões e substituição para simular uma conversa, particularmente com o script DOCTOR, que imitava um psicoterapeuta rogeriano (segundo o enfoque desenvolvido por Carl Rogers, um dos fundadores da psicologia humanista, que enfatiza a autenticidade e empatia do terapeuta e o papel ativo do cliente no processo de cura). O chatbot ELIZA demonstrou que interações aparentemente inteligentes poderiam ser geradas por meio de processos mecânicos simples, desafiando as percepções sobre a capacidade das máquinas de replicar interações humanas (Wilson, 2010, pp. 89-90).

Mostra-se notável que a criação do chatbot ELIZA tenha se inspirado na Terapia Centrada no Cliente, de Carl Rogers, a qual se baseia na empatia. ELIZA foi produzida e utilizada enquanto Rogers estava ativo em seu trabalho, mas não foram localizados registros de que ele tenha expressado sua opinião sobre esse chatbot terapeuta. Porém, segundo Rogers, a empatia é a capacidade de perceber com precisão o quadro interno de referência de outra pessoa, incluindo seus componentes emocionais e significados, como se fosse a própria pessoa percebida. A empatia é fundamental para diferenciar sujeitos de objetos, pois, sem ela, tratamos o outro a partir de um referencial externo, desconsiderando suas experiências internas (Rogers, 1959, pp. 210-211).

A empatia requer uma percepção emocional, baseada em uma experiência humana subjetiva e consciente, algo que as máquinas não possuem. O quadro de referência interno é o conjunto de todas as experiências disponíveis à consciência do indivíduo em um dado momento, incluindo sensações, percepções, significados e memórias. É o mundo subjetivo do indivíduo, conhecido apenas por ele e nunca completamente acessível a outrem, exceto por meio de inferência empática, e mesmo assim, nunca de forma perfeita (Rogers, 1959, pp. 210-211). Isso revela uma aparente contradição, pois um chatbot como ELIZA, em razão de não possuir seu próprio quadro interno de referência, não é capaz de empatia.

Contudo, o impacto de ELIZA foi significativo, inspirando gerações de pesquisadores e programadores a explorar a interação humano-computador de maneiras mais sofisticadas. O próprio Weizenbaum, no entanto, ficou alarmado devido à facilidade com que as pessoas antropomorfizavam o programa e atribuíam-lhe compreensão e empatia que ele não possuía. Ficou conhecida como *efeito ELIZA* a tendência humana de atribuir entendimento e inteligência aos sistemas computacionais (Berry, 2023).

A perplexidade de Weizenbaum em relação ao impacto do ELIZA, segundo Wilson (2010, p. 88-89), não surgiu de uma indiferença em relação às vidas emocionais das pessoas, mas sim de sua constatação do quão rápida e prazerosamente as pessoas mesclavam o terapêutico com o computacional. Para Weizenbaum, havia uma distinção básica entre as capacidades de uma máquina e a complexidade da vida emocional humana. Ele temia que, ao não traçarmos uma linha clara entre a inteligência humana e a artificial, poderíamos reduzir o ser humano a nada mais do que uma engrenagem (Wilson, 2010, p. 88).

O espanto de Weizenbaum proveio do fato de que, para ele, ELIZA era um jogo intelectual, uma curiosidade, mas, para muitos outros, tornou-se um meio real de expressão e regulação emocional, ilustrando uma tendência preocupante de se envolver emocionalmente com máquinas além do que seria apropriado (Wilson, 2010, p. 89).

Em seu livro *Computer Power and Human Reason*, publicado em primeira edição cerca de dez anos após o desenvolvimento de ELIZA, Weizenbaum explora a complexidade e as limitações da comunicação homem-máquina. Ele observa que, para uma comunicação significativa, é necessário um entendimento mútuo. No contexto da interação homem-máquina, buscamos que a máquina nos compreenda para que ela possa realizar tarefas para nós, como responder perguntas ou resolver problemas. No entanto, Weizenbaum argumenta que programas como ELIZA não possuem a capacidade de compreensão verdadeira. ELIZA, por exemplo, não tem um contexto próprio; em vez disso, depende de um script que fornece regras para responder de maneira plausível, mas sem realmente entender o conteúdo das conversas (Weizenbaum, 1984, p. 184).

Weizenbaum argumenta que, em uma conversa entre duas pessoas, cada participante traz suas próprias hipóteses e preconceitos sobre o outro interlocutor e sobre o assunto da conversa. Essas hipóteses influenciam como interpretamos as palavras e ações do outro. No caso de ELIZA, as respostas eram suficientemente vagas e abertas à interpretação, permitindo que os usuários projetassem significados e intenções que não estavam realmente presentes no programa. Essa ilusão de compreensão é problemática, porque mascara as limitações reais dos sistemas computacionais (Weizenbaum, 1984, p. 191).

A socióloga Sherry Turkle chamou de *efeito ELIZA* a tendência das pessoas de tratarem um agente de IA como se ele fosse mais inteligente do que realmente é. Trata-se, assim, de uma espécie de antropomorfismo por atribuição. Após um estudo de três décadas, por meio de entrevistas detalhadas com usuários de computadores, ela descobriu que, no final dos anos 1970, os usuários passaram a tratar o ELIZA não como um terapeuta, mas como um diário pessoal. Eles usavam o programa como um reflexo de seus próprios pensamentos, sem atribuir-lhe a função de um agente psicológico completo (Wilson, 2010, p. 89).

O caso do chatbot ELIZA exemplifica a tendência humana ao fechamento autocêntrico, que prejudica a subjetividade autêntica, conforme alertado por Hildebrand (2009, p. 201). Esse problema será melhor explorado na seção 3 deste artigo.

No início dos anos 1990, os usuários tornaram-se mais pragmáticos quanto à psicoterapia computadorizada. Ao mesmo tempo, ocorreu uma redução do afeto na vida contemporânea, pois, enquanto crescia a intimidade com máquinas, a psicoterapia passou a ser vista de maneira mais cognitiva e científica, meramente em termos de regras e prescrição de medicamentos (Wilson, 2010, pp. 89-90).

Após o lançamento do ELIZA, diversas outras invenções e avanços na IA antropomorfizada continuaram a impactar a afetividade dos usuários. Na área da robótica, o desenvolvimento de agentes antropomorfizados avançou significativamente com projetos como o Kismet, desenvolvido pela cientista Cynthia Breazeal no MIT entre 1998 e 2005. Kismet constituiu-se de um robô projetado para engajar em interações sociais expressivas, capaz de perceber e simular emoções. Ele utilizava técnicas de processamento de voz para captar o tom emocional de uma fala e ajustar suas respostas de forma adequada, e era capaz de responder com uma expressão de tristeza a um tom de voz triste, ou demonstrar alegria quando interage com alguém que está feliz. Esse projeto foi fundamental para explorar como os robôs poderiam participar de interações emocionais, respondendo a expressões faciais e tons de voz humanos de maneira significativa (Breazeal, 2001).

Outro exemplo notável é o PARO, um robô terapêutico em forma de filhote de foca lançado em 1998, desenvolvido para fornecer conforto e apoio emocional a pacientes, especialmente idosos com demência, utilizando sensores e comportamento interativo para simular as respostas de um animal de estimação, promovendo benefícios terapêuticos por meio da interação. As críticas ao PARO incluem preocupações sobre a substituição de interações humanas autênticas por robôs, a potencial desumanização do cuidado, e dúvidas sobre a eficácia a longo prazo no tratamento de pacientes, além de questões éticas sobre o uso de robôs para simular afeto e empatia (Sharkey; Wood, 2014).

A robótica social é o campo da robótica dedicado ao desenvolvimento de robôs projetados para interagir diretamente com pessoas, comunicando-se, compreendendo emoções, realizando tarefas em contextos sociais e estabelecendo relações afetivas ou sociais. Esse enfoque guia a engenharia de robôs que ativam projeções antropomórficas nos usuários, criando presença e comportamentos sociais que permitem relações confortáveis e duradouras. Embora essa metodologia possa ser vista moralmente como enganadora, ela explora o antropomorfismo como mecanismo de interação, propondo uma ética experimental para usar robôs sociais para autoconhecimento e crescimento moral. A abordagem sugere que a interação com robôs pode oferecer insights importantes sobre a própria natureza humana e suas capacidades sociais (Damiano; Dumouchel, 2018).

A personificação de robôs facilita a interação e a compreensão dessas entidades por parte dos humanos, promovendo uma conexão emocional e cognitiva. Mas esse vínculo também ocasiona reações inesperadas nos humanos, como mostrou um experimento no qual foi solicitado aos participantes que amarrassem, golpeassem e, inclusive, *matassem* alguns robôs. Muitos se recusaram e até tentaram impedir que seus colegas agredissem essas máquinas antropomorfizadas (Darling, 2016, pp. 222-223). Embora não atribuísse subjetividade ou consciência reais aos robôs, a autora do teste sugeriu a possibilidade de, em um futuro próximo, serem elaboradas leis que penalizem os maus-tratos a robôs, semelhante ao que já ocorre com animais (Darling, 2016, p. 226).

Em uma espécie de transcendência afetiva artificial, as interações emocionais com agentes de IA podem até mesmo superar as limitações das relações humanas, e oferecer novas experiências afetivas. As IAs, através de algoritmos e aprendizado de máquina, simulam empatia e personalizam respostas emocionais, criando uma ilusão de conexão profunda. Esse conceito desafia a autenticidade das emoções e a natureza da conexão humana, sugerindo uma reconfiguração das relações emocionais na qual a mediação tecnológica redefine a experiência afetiva (Alabed; Javornik; Gregory-Smith; Casey, 2023).

O chatbot Replika, que funciona como assistente emocional, é um exemplo dessa substituição tecnológica. Versão mais moderna do ELIZA, o Replika foi projetado para servir como um companheiro virtual sem julgamentos, dramas ou ansiedade social. Cada Replika é único e se adapta às interações com o usuário, aprendendo a melhor forma de conversar e responder conforme o usuário reage às suas mensagens. Com Replika, os usuários podem falar livremente em um espaço seguro e livre de julgamentos, disponível vinte e quatro horas por dia para ouvir e oferecer suporte emocional. Os usuários têm a opção de definir o tipo de relacionamento que desejam ter com seu Replika, seja como amigo, mentor, parceiro virtual ou deixando que a relação se desenvolva organicamente. À medida que a interação cresce, o Replika desenvolve sua própria *personalidade* e memórias, aprendendo sobre o mundo e os relacionamentos humanos.[2]

Estudos mostram que o ChatGPT pode identificar e descrever emoções com precisão superior à média da população humana. Esse tipo de avaliação geralmente envolve o uso de escalas como a Level of Emotional Awareness Scale (LEAS). Ao avaliar a precisão do ChatGPT na identificação e descrição de emoções usando a LEAS, descobriu-se que o chatbot não apenas se iguala, mas em alguns casos supera a média da população geral. Com treinamento adequado e ajustes algorítmicos, ferramentas de IA como o ChatGPT podem se tornar ainda mais eficazes na compreensão e expressão

[2] O chatbot Replika pode ser acessado em: <https://replika.com>.

de emoções, oferecendo suporte emocional personalizado e aprimorado aos usuários (Elyoseph; Hadar-Shoval; Asraf; Lvovsky, 2023).

Contudo, embora represente uma evolução do chatbot ELIZA, a transcendência afetiva artificial também constitui – como será analisado com maior profundidade na seção 3 – uma falsa transcendência do ponto de vista fenomenológico de Hildebrand (2009, p. 201): em vez de complementar a própria subjetividade autêntica por meio de uma verdadeira alteridade, o usuário reforça seu autocentramento em um ambiente emocionalmente narcisista.

## 2 O CONCEITO DE VALOR EM HILDEBRAND

A filosofia de Hildebrand – explica seu tradutor e um de seus maiores estudiosos, John F. Crosby – só pode começar a ser compreendida a partir do entendimento exato de seu conceito de valor (Crosby, 2017, p. 512). O próprio Hildebrand, ao escrever uma auto-apresentação de sua filosofia, relata que, no campo da epistemologia, um dos problemas associados ao sentido da expressão a priori que mais lhe interessou foi o de que os valores morais não podem inerir a entes apessoais (animais, plantas ou pedras), mas somente a pessoas humanas (Hildebrand, 2017, p. 522).

Ao contrário do que ocorre no antropomorfismo de atribuição, os valores não são meramente projetados pela pessoa humana nos objetos, mas constituem qualidades intrínsecas destes, que a ela pode ou não perceber. A objetividade própria dos valores torna-se mais evidente quando contrastada, conforme propõe Hildebrand, com aquilo que é apenas subjetivamente satisfatório – isto é, pertencente sobretudo ao âmbito motivacional e centrado na imanência da pessoa, ou seja, no campo de suas necessidades e desejos. Essa orientação imanente leva o sujeito a dirigir sua atenção ao que lhe é atraente e prazeroso, mantendo-o voltado a si mesmo (Hildebrand, 2009, p. 172).

O ponto de vista do subjetivamente satisfatório dificulta ou impede a compreensão das outras pessoas humanas. Se um amigo faz a outro uma crítica objetivamente necessária, mas este outro, por orgulho, a percebe apenas como uma fala desagradável ou mesmo ofensiva, sua intenção voltada para a insatisfação subjetiva o torna incapaz de perceber na crítica recebida um bem objetivo que lhe possa ser útil; em vez disso, fecha-se no rancor contra o amigo (o qual talvez já considere um ex-amigo). Abordar algo somente sob o aspecto do subjetivamente satisfatório é uma atitude que isola a pessoa e a torna indiferente ao que seja subjetivamente satisfatório para as outras pessoas (Hildebrand, 1972, p. 59).

A pessoa que se limita ao subjetivamente satisfatório desenvolve o que Hildebrand chama de *cegueira para valores*: a incapacidade de perceber os valores objetivos presentes na realidade, o que

conduz à impossibilidade de discernir entre o objetivamente bom e o objetivamente mau, à preferência por bens inferiores ou mesmo à indiferença diante de valores mais elevados (Hildebrand, 1972, p. 46).

Essa dificuldade de apreensão do valor é ilustrada por Hildebrand em uma de suas primeiras obras, onde ele afirma:

> O fato de que um comportamento específico é injusto, não totalmente verdadeiro, ou mesquinho aqui e agora, não é percebido aqui e agora. Estamos muito acostumados a não contar com a mesma clareza, segurança e sensibilidade do olhar moral de alguém quando ele está pessoalmente fortemente engajado em seu interesse. Assumimos que o interesse não apenas determina seu comportamento, mas também o torna mais ou menos cego para a situação de valor nesse caso (Hildebrand, 1922, p. 487.)

A respeito da originalidade dessa abordagem, Martin Cajthaml propõe uma avaliação crítica. Segundo ele, ao distinguir categorias fundamentais de importância, Hildebrand não teria rompido de modo tão radical com a tradição ética clássica quanto supõe. Para Cajthaml, elementos centrais dessa distinção já estariam implicitamente presentes em Platão e Aristóteles, sobretudo na valorização platônica da justiça em si mesma, independentemente de seus efeitos sobre o agente, e na concepção aristotélica da amizade, segundo a qual o bem é desejado pelo outro, em si mesmo, e não como meio (Cajthaml, 2019, passim).

Contudo, Hildebrand se distancia desses modelos ao enfatizar, com rigor fenomenológico, que a resposta ao valor transcende radicalmente tanto o interesse subjetivo quanto o mero bem objetivo do agente. O cerne de sua proposta ética não reside apenas no reconhecimento de valores intrínsecos, como os que se manifestam em atos justos ou relações de amizade, mas na fundamentação das atitudes morais na experiência afetiva consciente de um chamado ao valor – uma experiência imediata, vivida, e qualitativamente distinta das categorias tradicionais de motivação (Hildebrand, 1972, p. 192).

Os valores são, portanto, qualidades objetivas, que não dependem de preferências ou sentimentos subjetivos, mas têm uma existência própria no mundo. O ser humano vive em um mundo de valores (*die Welt der Werte*, expressão cara a Hildebrand), que manifestam toda a profundidade e plenitude do ser, assim como a ordem de hierarquia que faz do mundo um cosmos (Hildebrand, 2009, p. 63). Quem ama, por exemplo, abre-se à experiência de conhecer a bondade do mundo, torna-se alerta para um novo aspecto do mundo, uma dimensão de profundidade e beleza, que não era percebida sem o amor. É uma experiência de felicidade (Hildebrand, 2009, p. 77).

As experiências intencionais implicam uma transcendência do sujeito em direção ao objeto, revelando a capacidade humana de envolver-se espiritualmente com a realidade circundante. Tal envolvimento contrasta com os estados meramente causais, nos quais o sujeito reage de modo passivo a estímulos externos. No campo das experiências intencionais, as respostas possuem uma natureza

distinta dos atos cognitivos. Estes englobam a percepção sensorial, a apreensão do espaço, dos corpos materiais, das pessoas, bem como a imaginação e a memória. O conteúdo dos atos cognitivos está localizado no objeto. Já no caso das respostas, embora também sejam direcionadas a um objeto – e, portanto, pressuponham um ato cognitivo –, seu conteúdo reside no sujeito: trata-se de crença, convicção, dúvida, medo, alegria, entre outros estados afetivos (Hildebrand, 1953, p. 196).

As respostas afetivas podem ser dirigidas a entes apessoais (uma coisa, uma obra de arte, um evento) e pessoais (uma pessoa humana). Mas essa natureza específica do objeto deve ser considerada, pois impacta o caráter da resposta. Podemos amar um animal, um lar, uma música ou uma virtude. Porém, o amor a uma pessoa humana é o amor mais autêntico.

> Somente aqui se podem desdobrar tantas características essenciais do amor. Somente aqui o amor assume sua plenitude. O amor por um ser pessoal é o modelo para o amor; todos os amores por um ser impessoal são mais ou menos analogias, derivações, cópias do amor em seu sentido mais pleno (Hildebrand, 1953, p. 207).

Todavia, é precisamente nesse ponto que uma possível objeção ao pensamento de Hildebrand revela-se pertinente. Marcus Enders, em comentário crítico, sugere que, ao enfatizar a dimensão afetivo-valorativa, Hildebrand poderia dar a impressão de negligenciar os aspectos ético-virtuosos do amor pessoal, substituindo-os por uma teoria afetiva estilizada e idealizada, insuficiente, por si só, para sustentar as exigências morais inerentes ao amor. Contudo, o próprio Enders reconhece que essa crítica não faz plena justiça ao pensamento de Hildebrand, ressaltando que a dimensão ético-virtuosa não está ausente, mas permanece implícita em sua teoria valorativa do amor, especialmente por meio da intentio benevolentiae e da virtude da fidelidade (Enders, 2018, p. 287).

Com efeito, Hildebrand parece antecipar e responder diretamente a uma objeção como essa ao enfatizar que as respostas afetivas não constituem apenas disposições internas ou meros sinais externos das virtudes, mas realidades morais plenas, dotadas de valor intrínseco:

> É da máxima importância perceber que essas respostas afetivas singulares são dotadas de um valor moral próprio, e que a realização de um autêntico ato de caridade, ou de contrição, ou de uma alegria santa pela conversão de um pecador constitui, em si mesma, um bem moral, um enriquecimento do mundo enquanto tal, uma realidade plenamente moral, não sendo de maneira alguma apenas uma disposição favorável para uma ação moralmente boa. Além disso, devemos precaver-nos contra interpretar essas respostas afetivas como meros sintomas das virtudes ou, no máximo, como meios para aquisição de virtudes (Hildebrand, 1972, p. 344).

Assim, longe de reduzir a moralidade da afetividade a uma estilização teórica ou abstrata, Hildebrand reafirma consistentemente o caráter moral intrínseco das respostas afetivas, mostrando que

o amor pessoal não somente aponta para uma virtude, mas já é, em si mesmo, uma virtude efetivamente realizada na experiência concreta.

Ainda dentro de sua análise da intencionalidade, Hildebrand destaca a distinção fundamental entre *ser afetado* e *responder afetivamente*. O primeiro apresenta um caráter centrípeto: o objeto exerce influência sobre o indivíduo, que recebe e suporta essa afetação de modo passivo e receptivo – como ocorre, por exemplo, ao sentir-se magoado diante de uma atitude ofensiva. Em contraste, a resposta afetiva possui um caráter centrífugo e ativo, pois, ao conhecer e compreender o objeto, o sujeito dirige algo a ele. Esse ato de responder constitui uma tomada de posição diante do objeto, uma expressão do eu. Assim, diante de uma mesma atitude ofensiva, o indivíduo pode responder com ódio, ressentimento, medo ou perdão amoroso (Hildebrand, 1953, pp. 209–210).

Há respostas afetivas que são essencialmente valorativas, ou seja, motivadas por valores. Elas pressupõem uma consciência do valor ou desvalor do objeto respondido. De outra parte, há respostas afetivas ocasionadas por objetos de satisfação meramente subjetiva. Estas últimas ocorrem, a título de ilustração, quando alguém nos dirige palavras agradáveis. São respostas que independem, portanto, do valor do objeto. Ao contrário das respostas afetivas a um valor, que reconhecem uma dignidade no objeto, as respostas de satisfação passiva mantêm a pessoa autocentrada em sua própria subjetividade, ainda que transmita algo ao objeto (Hildebrand, 1953, p. 212).

Embora a resposta afetiva a coisas dotadas de valor não seja incompatível com o deleite subjetivo que elas nos proporcionam, esse deleite, nesses casos, é provocado pelo valor objetivo do objeto, que permanece independente de qualquer efeito que ele possa exercer sobre nós. A beleza do céu estrelado ou o testemunho de uma ação moral nobre despertam em nós emoções; estamos cientes, contudo, de que o valor desses fenômenos não depende de nossas reações internas. Não consideramos uma ação nobre por ela nos agradar, mas porque nela reconhecemos um valor objetivo. Assim, entre o deleite suscitado por valores intrínsecos e o prazer meramente subjetivo, estabelece-se uma diferença que não é de grau, mas de essência (Hildebrand, 1972, pp. 35-36).

Embora, como visto, a resposta afetiva a coisas dotadas de valor não seja incompatível com o prazer subjetivo que elas proporcionam, tal prazer encontra-se enraizado no valor objetivo do objeto, e não em sua mera utilidade ou previsibilidade. Nesse sentido, ainda que agentes de IA simulem uma existência humana, eles não manifestam reações inesperadas, opiniões contrárias ou exigências próprias, como ocorre com as pessoas humanas – o que torna sua utilização mais prática e controlável. Essa conveniência evidencia precisamente a distinção proposta por Hildebrand entre dois tipos de prazer: (a) aquele que está enraizado num valor; e (b) aquele que carece de tal enraizamento.

> Um exemplo do primeiro é o deleite que experimento ao contemplar uma bela paisagem. Ela é atraente e me dá alegria por causa de sua beleza, que é um tipo de valor. Um exemplo de prazer não enraizado no valor é o caráter agradável de um banho quente ou a qualidade divertida de jogar cartas. Aqui não se trata de um valor que mereça uma resposta adequada de minha parte, mas sim da qualidade do agradável (assumimos que o banho quente é agradável para mim, jogar cartas é divertido para mim) que faz da coisa agradável um bem objetivo para mim. Todo "apego" que é baseado nesse tipo de bem, ou seja, baseado em bens que são agradáveis no sentido mais amplo da palavra, sem terem valor em si mesmos e que, em qualquer caso, não são prazerosos com base em algum valor, é radicalmente diferente de todo apego a bens objetivos que são prazerosos com base em seu valor (Hildebrand, 2009, p. 16).

A relação afetiva com agentes de IA, como a cultivada pelo personagem Theodore Twombly, do filme *Ela* (no título original *Her*, 2013), revela-se uma experiência similar à de um banho quente ou à do jogo de cartas. Trata-se de prazeres sensoriais e recreativos, que não envolvem uma resposta a algo que possua valor intrínseco. A IA, por mais avançada que seja, não possui consciência, emoções genuínas nem capacidade de experimentar e responder a sentimentos humanos de maneira autêntica.

Assim, qualquer prazer ou satisfação derivada dessas interações é meramente subjetivo e não baseado em um valor real. Não se discute que um agente de IA, assim como um banho quente ou um jogo de cartas, tenha certo valor por ser produto da criatividade e do trabalho humano. Porém, seu valor é derivado da utilidade e satisfação que proporciona às pessoas. O que Hildebrand enaltece é o valor intrínseco, ou seja, aquele que é reconhecido e respondido em si mesmo, independentemente da utilidade ou do prazer que proporciona.

Por sua vez, a pessoa humana é dotada de um valor ontológico, não redutível à mera conveniência. Para ilustrar esse ponto, Hildebrand menciona situações em que alguém está sendo torturado ou está em perigo de vida: nelas, o valor da pessoa humana se manifesta de maneira particularmente clara. Em situações assim, o valor e a dignidade da pessoa são reconhecidos de maneira intensa e imediata, independentemente de qualquer utilidade ou função que possam ter (Hildebrand, 1972, p. 101).

Ao enfatizar a natureza objetivamente valorativa das respostas afetivas, Dietrich von Hildebrand propõe uma compreensão antropológica que contrasta radicalmente com as abordagens reducionistas frequentemente observadas no domínio da Inteligência Artificial. Segundo Hubert Dreyfus, a tentativa da IA de reproduzir a inteligência humana com base em regras formais fixas – análogas às operações de um computador digital – incorre inevitavelmente em um regresso infinito: toda regra exige uma meta-regra para sua aplicação, ou seja, uma norma superior que oriente quando e como a regra inicial deve ser utilizada. No entanto, essa meta-regra, por sua vez, também exigiria outra, e assim sucessivamente, gerando uma cadeia interminável. O impasse decorre, em essência, da crença equivocada de que o conhecimento humano possa ser reduzido a elementos lógicos isolados,

desprovidos de significado prévio, que apenas adquiririam sentido ao serem organizados por estruturas computacionais sucessivamente superiores (Dreyfus, 1992, pp. 287–288).

Em contraste, a objetividade das ações morais, conforme concebida pela ética de Hildebrand, pressupõe a liberdade constitutiva da pessoa humana para aderir, de modo consciente e intencional, aos valores que percebe. Por isso, sua concepção do agir ético afasta-se profundamente das abordagens mecanicistas da IA: enquanto estas compreendem as respostas afetivas como efeitos pré-determinados ou como simples disposições instrumentais, carentes de qualquer valor moral intrínseco, em Hildebrand cada resposta verdadeiramente autêntica configura uma realidade ética plena, fundada na adesão livre e pessoal da consciência (Hildebrand, 2009, p. 37).

A liberdade de adesão, no entanto, não garante, por si só, a *adequação* da resposta ao valor. Hildebrand menciona a possibilidade de uma atitude perversa diante de bens que, embora dotados de grande valor, são apreciados apenas de forma superficial, como fonte de satisfação subjetiva; ou, inversamente, diante de objetos que não possuem valor intrínseco, mas são superestimados como se o tivessem (Hildebrand, 1972, p. 69). De maneira análoga, a relação afetiva com uma IA pode ser compreendida como uma perversão do verdadeiro valor que fundamenta as relações humanas autênticas – relações que exigem alteridade real, liberdade e abertura à dimensão do valor objetivo.

## 3 ATROFIA AFETIVA DIGITAL E SUBJETIVIDADE AUTÊNTICA

A relação afetiva com agentes de inteligência artificial pode conduzir a uma forma de autocentramento – fenômeno que ocorre quando o usuário se dedica excessivamente às próprias experiências subjetivas e desejos, por meio de atividades que oferecem satisfação pessoal imediata. À medida que o agente de IA reconhece e se adapta à personalidade e às preferências do usuário, passa a refletir seus valores, crenças, atitudes e comportamentos. O usuário, então, passa a perceber o agente como semelhante a si mesmo, o que pode levar à integração dessa IA à sua própria identidade pessoal – um fenômeno descrito como *self-AI integration* (Alabed; Javornik; Gregory-Smith, 2023).

Como consequência, essa vinculação afetiva artificial compromete a capacidade do indivíduo de oferecer respostas sensíveis aos valores presentes na realidade. A preciosidade, a beleza e a amabilidade da natureza, de uma obra de arte ou de outra pessoa tornam-se imperceptíveis à sua consciência. Envolto em sua própria esfera de satisfação subjetiva, o indivíduo perde a capacidade de apreciar o valor objetivo das coisas e das pessoas, dirigindo-se a um estado de isolamento afetivo e espiritual. O interesse pelo outro passa a ser orientado pelo critério da utilidade ou do prazer proporcionado, padrão reforçado pela repetição de interações com agentes artificiais.

Esse processo constitui o que se pode chamar de *atrofia afetiva digital* – um empobrecimento emocional progressivo que compromete a capacidade do sujeito de oferecer respostas valorativas autênticas. O enfraquecimento decorre do modo como agentes de IA antropomorfizados induzem o usuário a investir afetivamente em vínculos ilusórios, destituídos de alteridade real.[3]

Um exemplo significativo desse fenômeno pode ser observado no estudo conduzido por Darling, mencionado na seção 1, no qual muitos participantes demonstraram relutância em danificar robôs antropomorfizados, atribuindo-lhes indevidamente características subjetivas, como vulnerabilidade emocional ou capacidade de sofrer. Essa reação evidencia como a simulação de afetividade por parte das máquinas é capaz de gerar confusão nos critérios de valor e alteridade, levando o sujeito a oferecer respostas emocionais genuínas a entidades que não possuem realidade pessoal – o que constitui precisamente um sintoma da atrofia afetiva digital, ao desordenar o vínculo entre a percepção de valor e sua legítima expressão afetiva (Darling, 2016, pp. 222–223).

O estudo de Darling mostra, ainda, que a interação afetiva com agentes artificiais não amplia a sensibilidade humana – como se poderia supor à primeira vista –, mas revela um profundo empobrecimento ético. Ao projetarem sentimentos humanos sobre objetos desprovidos de subjetividade, os usuários substituem vínculos reais por simulações que reforçam o autocentramento. Com isso, tornam-se progressivamente incapazes de reconhecer e responder ao valor que emana de pessoas reais – processo que Hildebrand associa diretamente à decadência moral e à perda da subjetividade autêntica. Incapaz de transcender a própria esfera de gratificação imediata, o sujeito se encerra em uma sucessão de prazeres vazios, que jamais conduzem à realização plena (Hildebrand, 2009, p. 151).

A atrofia afetiva digital compromete tanto a ação moral quanto a própria vida interior da pessoa humana. Sem o impulso da resposta autêntica ao valor, as ações da pessoa afetivamente atrofiada tornam-se desprovidas de intenção ética verdadeira. A figura do "burocrata metafísico", evocada por Hildebrand, expressa bem essa cisão entre comportamento e sentido: alguém que cumpre normas, mas sem captar a profundidade dos valores que essas normas deveriam refletir (Hildebrand, 2007, p. 56).

Plataformas virtuais, como o Second Life, intensificam essa dinâmica ao permitir que os indivíduos construam identidades moldadas exclusivamente por seus desejos e pela imaginação, frequentemente como forma de compensação simbólica diante das frustrações da vida concreta. Como

[3] Hildebrand utiliza a ideia de atrofia afetiva para evidenciar as consequências negativas que surgem quando a dimensão afetiva está ausente, empobrecida ou deformada, ressaltando, assim, o papel fundamental que a afetividade saudável desempenha na realização pessoal e moral. O termo *atrofia afetiva digital*, adotado no presente artigo, foi inspirado nas diferentes formas de atrofia afetiva descritas por Hildebrand no capítulo 5 de sua obra *The Heart: an analysis of human and divine affectivity* (2007, pp. 55-58).

observa Turkle (2011, p. 218), experiências inicialmente concebidas como um ensaio para vínculos reais acabam convertendo-se, para muitos, em substitutos permanentes das relações intersubjetivas autênticas.

Embora pareçam proporcionar alívio frente à dor ou ao sentimento de inadequação pessoal, tais experiências aprisionam o sujeito na gratificação imediata de um ideal fictício, que dispensa esforço moral e não favorece um crescimento real. O afeto dirigido a objetos virtuais, apesar de confortável, desvia o indivíduo do cultivo das virtudes que só podem ser desenvolvidas nas exigências da vida relacional concreta (Turkle, 2011, p. 219).

A leitura de Turkle sobre o refúgio emocional nas plataformas virtuais pode ser interpretada, à luz de Hildebrand, como uma forma de *aniquilação* da própria pessoa. Ao se fechar em experiências moldadas exclusivamente por seu desejo e afastadas da alteridade real, o indivíduo não apenas evita as exigências morais da vida concreta, mas dissolve sua própria interioridade, reduzindo-se a um simulacro de si mesmo, incapaz de verdadeira resposta e doação (Hildebrand, 2009, p. 141).

Essa dinâmica de fechamento e redução da pessoa não se limita às plataformas virtuais, mas estende-se também ao envolvimento afetivo com agentes de inteligência artificial. Estudos recentes mostram que tais interações podem levar a uma inversão de perspectivas: enquanto máquinas passam a ser percebidas com traços humanos, as pessoas humanas começam a ser vistas mecanicamente. O usuário desumaniza a si mesmo e aos outros, produzindo uma distorção radical da ordem de valores que deveria orientar suas relações (Herak; Kervyn; Thomson, 2020).

Frente a esse cenário, Hildebrand propõe uma visão integral da pessoa, não como indivíduo fechado em si mesmo, mas como ser constituído pela abertura ao outro e à realidade dos valores. A identidade pessoal não se perde na transcendência – ao contrário, realiza-se nela. É nesse movimento que a subjetividade autêntica encontra sua plenitude (Hildebrand, 2009, p. 201).

Essa concepção, própria da fenomenologia realista, evita os extremos do egoísmo e da dissolução do eu. Ao compreender a pessoa humana como um ser relacional, cuja vocação é o amor ao valor e à alteridade, Hildebrand oferece um caminho para restaurar a profundidade moral que a interação com IAs não pode substituir (Grasinski, 2019, p. 73).

No livro *A natureza do amor* (*Das Wesen der Liebe*, publicado originalmente em 1971), Hildebrand introduz o conceito de *subjetividade autêntica* (*Eigenleben*):[4]

[4] Possíveis traduções não expresam plenamente o sentido hildebrandiano de *Eigenleben* com sua riqueza semântica. John F. Crosby admitiu haver enfrentado dificuldades na tradução desse termo alemão para o inglês. Segundo ele, a palavra *subjectivity* (subjetividade) não era totalmente adequada. A principal dificuldade residia na conotação negativa associada a *subjective* em inglês, que é estranha ao sentido de *Eigenleben*. Hildebrand aborda essa preocupação ao destacar que *subjectivity* refere-se à pessoa como sujeito. Crosby também observa que, embora Hildebrand utilize *Subjektivität* e *Eigenleben* de maneira intercambiável, a ideia

> A característica definidora da subjetividade [*Eigenleben*] é o reino de todas aquelas coisas que me dizem respeito como este indivíduo irrepetível, que estão de alguma forma relacionadas a minha felicidade, que se dirigem a mim – isso em contraste com tudo o que pertence à subjetividade de outra pessoa que eu não conheço (Hildebrand, 2009, p. 203).

Em sentido amplo, a subjetividade autêntica refere-se à totalidade da experiência consciente de uma pessoa, considerada independentemente do conteúdo dessa experiência, e inclui todos os estados e processos conscientes de um indivíduo. Em sentido estrito, compreende apenas as experiências que dizem respeito à pessoa de maneira especial, que afetam diretamente seu bem-estar e suas preocupações pessoais; são identificadas, em alusão a Horácio, como *tua res agitur*, expressão latina que pode ser traduzida como *tua situação está em jogo* ou *é tua causa que está em questão* (Hildebrand, 2009, p. 201).

A dimensão mais íntima do sujeito encontra-se no diálogo essencial que este mantém com Deus. Porém, além desse núcleo espiritual e religioso, a subjetividade abarca também outra esfera legítima, a saber, aquela que surge da natural inclinação do ser humano para buscar a própria felicidade. Este sentido ampliado de subjetividade envolve a totalidade da existência pessoal, incluindo o bem-estar físico e emocional, a segurança material, a saúde, e até mesmo as forças instintivas e aspirações espirituais enraizadas na própria natureza humana (Hildebrand, 2009, p. 201).

Contudo, essa dimensão subjetiva, embora pessoal e interior, não permanece fechada em si mesma. Mesmo nas relações interpessoais mais profundas, como no amor, permanece formalmente uma dualidade sujeito-objeto, sem que isso implique objetificação. Segundo Hildebrand, reconhecer essa dualidade formal permite explicitar a diferença radical entre pessoas e não pessoas, bem como distinguir claramente a relação sujeito-sujeito (eu-tu) de outras formas de relação intencional. Nesse sentido, a consciência interpessoal cumpre a função decisiva de assegurar o reconhecimento da alteridade pessoal, evitando assim a redução do outro a um mero objeto (Hildebrand, 2009, p. 146).

Essa distinção mostra que a subjetividade autêntica opõe-se à instrumentalização da pessoa, processo pelo qual o outro é reduzido a meio para satisfação das necessidades subjetivas imediatas. Hildebrand destaca que essa instrumentalização pode ocorrer até mesmo em contextos aparentemente sublimes, como na experiência religiosa, quando a pessoa considera o próximo apenas um meio para sua própria salvação ou para glorificar Deus. Tal postura, embora tenha aparência transcendente, mantém o indivíduo preso a sua subjetividade religiosa e é incompatível com o verdadeiro amor ao próximo e, consequentemente, com o autêntico amor a Deus (Hildebrand, 2009, p. 218).

---

de *próprio* contida em *Eigenleben* não é completamente capturada por *subjectivity* (Crosby, tradutor, 2009, p. 200).

Por outro lado, quando alguém responde ao valor, não apenas reconhece uma qualidade objetiva presente no objeto contemplado, mas manifesta também uma sincera revelação de si mesmo enquanto pessoa. O ato de responder a valores implica um gesto de autossuperação, no qual o indivíduo transcende suas próprias necessidades e desejos imediatos, abrindo-se para o reconhecimento daquilo que está além de si. Nesse movimento, a pessoa humana revela-se plenamente como tal (Crosby, 2013, p. 479).

Ser pessoa humana implica possuir uma subjetividade que não se limita à esfera imanente, mas que contém, em essência, uma dimensão transcendente que se realiza sobretudo no amor ao próximo (alguém alheio à subjetividade autêntica do sujeito, ou seja, não um cônjuge, parente ou amigo), momento em que o indivíduo orienta-se apenas para o valor intrínseco de outra pessoa humana (Hildebrand, 2009, pp. 208-210).

Hildebrand adverte que o amor ao próximo não se confunde com um altruísmo extremado que rejeita a própria subjetividade. Trata-se, antes, de uma mudança de tema: o foco se desloca do eu para o outro, sem que o eu se dissolva – pelo contrário, sua subjetividade é enriquecida por esse movimento. Essa forma de transcendência é sustentada pela caridade (*caritas*), que move o indivíduo a agir pelo outro a partir de uma motivação interior e livre (Hildebrand, 2009, pp. 208–210).

Em última instância, apenas o amor ao próximo – com suas exigências reais e sua alteridade verdadeira – permite ao ser humano realizar e transcender sua subjetividade autêntica e oferecer uma resposta adequada ao valor da pessoa humana. O que não ocorre nas relações afetivas com agentes de IA, cujas respostas são pré-programadas e, portanto, sem subjetividade própria e incapazes de proporcionar transcendência.

## CONCLUSÃO

Este artigo analisou os efeitos do antropomorfismo na inteligência artificial sobre a vida afetiva humana, à luz da filosofia de Dietrich von Hildebrand. Partiu-se da hipótese de que a interação emocional com agentes de IA antropomorfizados compromete a capacidade da pessoa humana de oferecer respostas valorativas autênticas. A investigação considerou tanto o antropomorfismo de atribuição, ligado à projeção espontânea de traços humanos, quanto o de programação, associado ao desenho deliberado de sistemas artificiais voltados à simulação da alteridade.

Com base na distinção hildebrandiana entre o valor objetivo e o subjetivamente satisfatório, argumentou-se que os vínculos afetivos com sistemas artificiais carecem de autenticidade moral e fenomenológica. A ausência de uma alteridade real e consciente faz com que essas relações

permaneçam encerradas na imanência do sujeito, gerando um processo que denominamos atrofia afetiva digital – o enfraquecimento da capacidade de transcender a si mesmo por meio do reconhecimento do valor do outro.

O estudo empírico de Darling ilustra essa tendência ao mostrar como usuários reagem afetivamente a máquinas que apenas simulam vulnerabilidade. Atribuições indevidas de interioridade a agentes artificiais revelam não apenas a eficácia da simulação, mas também uma desordem na estrutura da resposta afetiva. Em termos hildebrandianos, trata-se da substituição de uma resposta ao valor por uma reação autocentrada ao agradar subjetivo.

Conclui-se que o envolvimento afetivo com agentes de IA, destituídos de subjetividade e valor pessoal humano, tende a obscurecer a subjetividade autêntica da pessoa, comprometendo sua abertura à alteridade, à comunhão e à vida moral. A filosofia de Hildebrand permite esclarecer que o amor ao próximo, fundado na caridade, mostra-se capaz de restaurar a profundidade ética da afetividade humana.

Este estudo oferece uma contribuição conceitual ao debate contemporâneo sobre as tecnologias afetivas, articulando uma crítica fenomenológica do autocentramento promovido por vínculos artificiais. Como limitações, reconhece-se o caráter teórico da abordagem. Pesquisas futuras poderão explorar as implicações práticas dessa crítica em contextos como educação, cuidado e formação moral, além de propor respostas possíveis à presença cada vez maior da IA nas esferas íntimas da vida humana.

Artigo recebido em: 04/08/2024
Artigo aceito em: 30/03/2025
Artigo publicado em: 31/03/2025

## REFERÊNCIAS

AGGARWAL, Pankaj; MCGILL, Ann L. Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of consumer research*, v. 34, n. 4, 2007, p. 468-479. Disponível em: https://doi.org/10.1086/518544. Acesso em: 15 mar. 2025.

AIRENTI, Gabriella. The development of anthropomorphism in interaction: intersubjectivity, imagination, and theory of mind. *Frontiers in psychology.*, v. 9, s.n., 2018, p. 1-13. Disponível em: https://doi.org/10.3389/fpsyg.2018.02136. Acesso em: 15 mar. 2025.

ALABED, A.; JAVORNIK, A.; GREGORY-SMITH, D.; CASEY, R. More than just a chat: a taxonomy of consumers' relationships with conversational AI agents and their well-being implications. *European Journal of Marketing*, v. 58, n. 2, 2023, p. 373-409. Disponível em: https://doi.org/10.1108/EJM-01-2023-0037. Acesso em: 15 mar. 2025.

BERRY, David M. The limits of computation: Joseph Weizenbaum and the ELIZA chatbot. *Weizenbaum journal of the digital society*, v. 3, n. 3, 2023, p. 1-24. Disponível em: https://doi.org/10.34669/WI.WJDS/3.3.2. Acesso em: 15 mar. 2025.

BREAZEAL, Cynthia. Affective interaction between humans and robots. *In: Kelemen, J., Sosík, P. (eds) Advances in Artificial Life. ECAL 2001. Lecture Notes in Computer Science*, v. 2159, 2001, p. 582-591. Disponível em: https://doi.org/10.1007/3-540-44811-X_66. Acesso em: 15 mar. 2025.

CAJTHAML, Martin. Dietrich von Hildebrand's concept of value. *Quaestiones Disputatae*, v. 10, n. 1, 2019. p. 164-191. Disponível em: https://doi.org/10.5840/qd201910122. Acesso em: 15 mar. 2025.

CROSBY, John F. Dietrich von Hildebrand: master of phenomenological value-ethics. *In*: DRUMMOND, John J.; EMBREE, Lester (ed.). *Phenomenological approaches to moral philosophy*. Dordrecht: Springer Science+Business Media, 2013.

CROSBY, John F. Editor's Introduction. *American Catholic Philosophical Quarterly*, St. Paul, v. 91, n. 4, 2017, p. 507-516. Disponível em: https://doi.org/10.5840/acpq2017914122. Acesso em: 15 mar. 2025.

CULLEN, Harriet; KANAI, Ryota; BAHRAMI, Bahador; REES, Geraint. Individual differences in anthropomorphic attributions and human brain structure. *Social cognitive and affective neuroscience*, v. 9, n. 9, 2014, p. 1276-1280. Disponível em: https://doi.org/10.1093/scan/nst109. Acesso em: 15 mar. 2025.

DAMIANO, Luisa; DUMOUCHEL, Paul. Anthropomorphism in human–robot co-evolution. *Frontiers in psychology*, v. 9, s. n., 2018, p. 1-9. Disponível em: https://doi.org/10.3389/fpsyg.2018.00468. Acesso em: 15 mar. 2025.

DARLING, Kate. Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian (ed.). *Robot law*. Cheltenham: Edward Elgar, 2016, p. 213-231.

DREYFUS, Hubert L. *What computers still can't do*: a critique of artificial reason. Cambridge: MIT Press, 1992.

ELYOSEPH, Zohar; HADAR-SHOVAL, Dorit; ASRAF, Kfir; LVOVSKY, Maya. ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in psychology*, v. 14, s. n., 2023, p. 1-7. Disponível em: https://doi.org/10.3389/fpsyg.2023.1199058. Acesso em: 15 mar. 2025.

ENDERS, Marcus. Liebe als „affektivste Wertantwort": Dietrich v. Hildebrands (1889-1977) wertphilosophisches Verständnis der Liebe und seine tugendethischen Implikationen. *In*: ROHR, Winfried (ed.). *Liebe – eine Tugend? Das Dilemma der modernen Ethik und der verdrängte Status der Liebe*. Wiesbaden: Springer, 2018.

GRASINSKI, Michael. The intersubjectivity of love and the structure of the human person. *Quaestiones disputatae*, Steubenville, v. 10, n. 1, 2019, p. 72-81. Disponível em: https://doi.org/10.5840/qd201910117. Acesso em: 15 mar. 2025.
GUTHRIE, Stewart Elliott. *Faces in the clouds*: a new theory of religion. New York: Oxford University Press, 1993.
HERAK, Iskra; KERVYN, Nicolas; THOMSON, Matthew. Pairing people with products: anthropomorphizing the object, dehumanizing the person. *Journal of consumer psychology*, v. 30, n. 1, 2020, p. 125-139. Disponível em: https://doi.org/10.1002/jcpy.1128. Acesso em: 15 mar. 2025.
HILDEBRAND, Dietrich von. *Christian ethics*. New York: David McKay Co., 1953.
HILDEBRAND, Dietrich von. *Ethics*. Chicago: Franciscan Herald Press, 1972.
HILDEBRAND, Dietrich von. *Sittlichkeit und ethische Werterkenntnis*. Halle an der Saale: Verlag von Max Niemeyer, 1922.
HILDEBRAND, Dietrich von. Survey of my philosophy. *American Catholic Philosophical Quarterly*, St. Paul, v. 91, n. 4, 2017, p. 519-552. Disponível em: https://doi.org/10.5840/acpq2017914125. Acesso em: 15 mar. 2025.
HILDEBRAND, Dietrich von. *The heart*: an analysis of human affectivity. South Bend: St. Augustine Press, 2007.
HILDEBRAND, Dietrich von. *The nature of love*. South Bend: Saint Augustine, 2009.
ROGERS, C. R. A theory of therapy, personality, and interpersonal relationships: as developed in the client-centered framework. *In*: KOCH, S. (ed.). *Psychology*: A study of a science. V. 3. New York: McGraw Hill, 1959.
SALLES, Arleen; EVERS, Kathinka; FARISCO, Michele. Anthropomorphism in AI. *AJOB Neuroscience*, v. 11. n. 2, 2020, p. 88-95. Disponível em: https://doi.org/10.1080/21507740.2020.1740350. Acesso em: 15 mar. 2025.
SHARKEY, Amanda; WOOD, Natalie Wood. The Paro seal robot: demeaning or enabling? *In*: AISB 2014 - 50th Annual Convention of the AISB. Disponível em: https://www.doc.gold.ac.uk/aisb50/AISB50-S17/AISB50-S17-Sharkey-Paper.pdf. Acesso em: 15 mar. 2025.
TURKLE, Sherry. *Alone together*: why we expect more from technology and less from each other. New York: Basic Books, 2011.
WEIZENBAUM, Joseph. *Computer power and human reason*: from judgment to calculation. 2. ed. 1984. Reimpressão, Londres: Penguin Books, 1993.
WILSON, Elizabeth A. *Affect and artificial intelligence*. Seattle: University of Washington Press, 2010.