

EFEITO ELIZA: IMPLICAÇÕES DA PROJEÇÃO ANTROPOMÓRFICA À INTEGRIDADE PSÍQUICA DIANTE DA UTILIZAÇÃO DE *CHATBOTS* PARA CUIDADOS COM A SAÚDE MENTAL

ELIZA EFFECT: IMPLICATIONS OF ANTHROPOMORPHIC PROJECTION ON PSYCHIC INTEGRITY IN THE FACE OF THE USE OF CHATBOTS FOR MENTAL HEALTH CARE

EFEECTO ELIZA: IMPLICACIONES DE LA PROYECCIÓN ANTROPOMORFICA PARA LA INTEGRIDAD PSÍQUICA RESPECTO AL USO DE CHATBOTS PARA EL CUIDADO DE LA SALUD MENTAL

Raissa Arantes Tobbin^{*}

Valeria Silva Galdino Cardin^{**}

1 Introdução. 2 *Chatbots* e a interação humano-máquina. 3 Dissonância cognitiva e projeção antropomórfica. 4 Utilização de *chatbots* no âmbito dos cuidados com a saúde mental: implicações ao direito da personalidade à integridade psíquica. 5 Conclusão. Referências.

RESUMO

Contextualização: Nos últimos anos, a utilização de *chatbots*, que podem ser conceituados como programas de computador que possuem habilidades para manter uma conversação com humanos, tem ganhado relevo na área da saúde, sobretudo para fins de cuidados com a saúde mental. Tais sistemas inteligentes, disponíveis para interação a qualquer tempo, podem conduzir exercícios para o controle de sintomas de estresse, ansiedade, depressão e ideação suicida. Diante do avanço tecnológico, é essencial verificar se as implicações da projeção antropomórfica, que é a tendência humana de atribuir

^{*} Doutoranda em Direito pela Universidade Cesumar (UNICESUMAR); Mestre em Ciências Jurídicas pela Universidade Cesumar (UNICESUMAR); Graduada em Direito pela Universidade Paranaense (UNIPAR); Graduada em Letras - Português/Espanhol pela Universidade Estadual de Ponta Grossa (UEPG); Advogada no Paraná, PR, BR. E-mail: <tobbinraissa@hotmail.com>. ORCID: <https://orcid.org/0000-0002-3655-8407>

^{**} Pós-Doutora em Direito pela Universidade de Lisboa; Doutora e mestre em Direito das Relações Sociais pela Pontifícia Universidade Católica de São Paulo (PUCSP); Docente da Universidade Estadual de Maringá e no Programa de Pós-Graduação em Ciências Jurídicas pela Universidade Cesumar (UNICESUMAR); Pesquisadora pelo ICETI; Advogada no Paraná. Maringá, PR, BR. E-mail: <valeria@galdino.adv.br>. ORCID: <https://orcid.org/0000-0001-9183-0672>



características humanas a objetos, em tais conversas, não poderiam ofender a integridade psíquica do indivíduo, um direito da personalidade.

Objetivo: O presente trabalho tem por objetivo analisar as implicações do Efeito Eliza e da projeção antropomórfica ao direito da personalidade à integridade psíquica diante da eventual utilização de *chatbots* para fins de cuidados com saúde mental.

Método: A pesquisa utilizou o método hipotético-dedutivo, fundamentado em revisão bibliográfica.

Resultados: Como resultado, verificou-se que há uma tendência humana de antropomorfizar dispositivos tecnológicos, lendo respostas como resultados de emoções humanas, o que pode provocar apego, identificação e dependência, diante da percepção de reciprocidade, cenário que pode gerar ofensa à integridade psíquica.

Conclusões: Há implicações éticas e jurídicas no que tange a sistemas de conversação para fins terapêuticos, especialmente em relação ao direito à integridade psíquica. É importante considerar também as questões ligadas à privacidade das informações compartilhadas, à segurança e a vieses discriminatórios.

Palavras-chave: direitos da personalidade; Efeito Eliza; saúde mental; *chatbots*; inteligência artificial.

ABSTRACT

Contextualization: In recent years, the use of chatbots, which can be define as computer programs that have the ability to hold conversations with humans, has gained relevance in the health field, especially for mental health care purposes. These intelligent systems, available for interaction at any time, can conduct exercises to control symptoms of stress, anxiety, depression and suicidal ideation. Despite such technological advances, it is essential to verify whether the implications of anthropomorphic projection, which is the human tendency of understand human characteristics of objects, in such conversations could not offend the individual's psychological integrity, a personality right.

Objective: the study aims to analyze the implication of the Eliza Effect and anthropomorphic projection in the right to personality right to psychic integrity in the face of the use of chatbots for mental health care purposes.

Method: the research used the hypothetical-deductive method, based on a bibliographic review.

Results: There is a human tendency to anthropomorphize technological devices, reading responses as results of human emotions, which can cause attachment, identification and dependence, given the perception of reciprocity, a scenario that can generate offense to psychic integrity.

Conclusions: There are ethical and legal implication related to conversation systems for therapeutic purposes, especially in relation to the right to psychic integrity. Issues related to the privacy of shared informatin, security and discriminatory biases are highlighted.

Keywords: personality rights; Eliza Effect; mental health; chatbots; artificial intelligence.

RESUMEN

Contextualización: En los últimos años, el uso de *chatbots*, que pueden conceptualizarse como programas informáticos que tienen la capacidad de mantener una conversación con humanos, ha ganado importancia en el sector salud, especialmente con fines de atención de la salud mental. Estos sistemas inteligentes, disponibles para interactuar en cualquier momento, pueden realizar ejercicios para controlar los síntomas de estrés, ansiedad, depresión e ideas suicidas. A pesar de tal avance tecnológico, es fundamental verificar si las implicaciones de la proyección antropomórfica, que es la tendencia humana a atribuir características humanas a los objetos, en tales conversaciones no podrían ofender la integridad psíquica del individuo, un derecho de la personalidad.

Objetivo: El presente Trabajo tiene como objetivo analizar las implicaciones del Efecto Eliza y la proyección antropomórfica sobre el derecho de la personalidad a la integridad psíquica ante el uso de chatbots con fines de atención de la salud mental.

Método: La investigación utilizó el método hipotético-deductivo, basándose en una revisión bibliográfica.

Resultados: Existe una tendencia humana a antropomorfizar los dispositivos tecnológicos, leyendo las respuestas como resultado de las emociones humanas, lo que puede provocar apego, identificación y dependencia, ante la percepción de reciprocidad, escenario que puede generar ofensa a integridad psíquica.

Conclusiones: Existen implicaciones éticas y legales a respecto de los sistemas de conversación con fines terapéuticos, especialmente en relación con el derecho a la integridad psicológica. Se destacan cuestiones relacionadas con la privacidad de la información compartida, la seguridad y el sesgo discriminatorio.

Palabras clave: derechos de la personalidad; Efecto Eliza; salud mental; *chatbots*; inteligencia artificial.

1 INTRODUÇÃO

Nas últimas décadas, tem crescido mundialmente a conscientização acerca da necessidade de cuidados com a saúde mental e da manutenção de um estilo de vida equilibrado e compatível com condições que proporcionem bem-estar físico e mental.

Segundo dados da Organização Mundial da Saúde (OMS) e da Organização Pan-Americana da Saúde (OPAS) (WHO, 2022b), sobretudo divulgados no “Relatório Mundial sobre Saúde Mental: Transformar a Saúde Mental para Todos”, em 2019, quase um bilhão de pessoas, incluindo 14% dos adolescentes do mundo, viviam com algum tipo de transtorno mental, sendo o suicídio responsável por uma em cada cem mortes. Os transtornos mentais são a principal causa de incapacidade, e as pessoas com condições graves de saúde mental morrem em média dez a vinte anos mais cedo do que a população em geral (WHO, 2022b, 2023).

A situação de crise sanitária ocasionada pela COVID-19 intensificou quadros de ansiedade, depressão e demais transtornos mentais. Durante o primeiro ano da pandemia, a depressão e a ansiedade aumentaram mais de 25%. Contudo, oito em cada dez pessoas com uma doença mental grave não receberam tratamento¹ (WHO, 2022b, 2023). Apesar de tais dados, verifica-se que os tratamentos e os cuidados com saúde mental ainda são considerados inacessíveis para boa parte da população mundial. Para contornar tal perspectiva, ganham relevo as soluções com base em *e-Health*, saúde móvel, telemedicina e sistemas de Inteligência Artificial (IA) baseados no tratamento de dados sobre saúde.

Os *chatbots* podem ser conceituados como programas de computador que possuem habilidades para manter uma conversação com humanos, sendo, a cada dia mais, ágeis e eficientes. Tais modelos, disponíveis para conversação a qualquer tempo, têm sido utilizados para experimentos com fins terapêuticos, como é o caso do *chatbot* Woebot, que conduz os usuários por meio de exercícios baseados na Terapia Cognitivo-Comportamental, sobretudo para fins de controle de sintomas de estresse, ansiedade, depressão e ideação suicida. Estima-se que a indústria de aplicação dos *chatbots* atinja US\$200 bilhões até 2026 (Brown, 2022; Han; Yin; Zhang, 2023).

Diante do avanço tecnológico, é essencial verificar se as implicações da projeção antropomórfica, que é a tendência humana de atribuir características humanas a objetos, em tais conversas não poderiam ofender a integridade psíquica do indivíduo, um direito da personalidade. Dessa forma, o presente trabalho tem por objetivo analisar as implicações do Efeito Eliza e da projeção antropomórfica à integridade psíquica, um direito da personalidade, diante da eventual utilização de *chatbots* para fins de cuidados com a saúde mental. O problema que orientou a condução da pesquisa pode ser sintetizado da seguinte forma: com base na análise do Efeito Eliza, a utilização da *chatbots* para fins terapêuticos poderia ofender o direito da personalidade à integridade psíquica?

A hipótese inicial é a de que a tendência humana de antropomorfizar dispositivos tecnológicos, lendo respostas de sistemas como fruto de emoções humanas, pode prejudicar ainda mais o desafiador trabalho terapêutico, que envolve cuidados com a

¹ Conforme o Relatório “Uma Nova Agenda para a Saúde Mental nas Américas”, de 2023, da OMS, quase um terço de todos os anos vividos com deficiência (YLD) e um quinto dos anos de vida ajustados por deficiência (DALYs) são causados por doenças mentais, doenças neurológicas, uso de substâncias e suicídio; os transtornos depressivos e de ansiedade são a terceira e a quarta principais causas de deficiência; o álcool é responsável por 5,5% de todas as mortes nas Américas; as Américas são a segunda região com maior consumo de álcool no mundo; o suicídio tira a vida de quase 100 mil pessoas por ano na região; a taxa regional de suicídio ajustada por idade aumentou em 17% entre 2000 e 2019; oito em cada dez pessoas com uma doença mental grave não recebem tratamento; em 2020, durante a pandemia da COVID-19, os transtornos depressivos graves aumentaram em 35% e os transtornos de ansiedade em 32%; 65% dos países relataram interrupções nos serviços essenciais de saúde mental e uso de substâncias em 2020. Esse índice caiu para 14% no início de 2023 (WHO, 2023). Ressalta-se, ainda, a existência do Plano de Ação Integral de Saúde Mental 2013-2030, por meio do qual os 194 países membros da OMS se comprometeram com metas globais para transformar a saúde mental (WHO, 2021).

saúde mental, com repercussão à integridade psíquica, um direito da personalidade. O estudo teve por objetivo geral analisar as implicações do Efeito Eliza e da projeção antropomórfica ao direito da personalidade à integridade psíquica diante da eventual utilização de *chatbots* para fins de cuidados com saúde mental.

Para alcançar o objetivo geral, foram traçados os seguintes objetivos específicos, que correspondem às seções de desenvolvimento do artigo:

- a) descrever a utilização de *chatbots* para fins de cuidados com a saúde mental, utilizando como exemplo o experimento com o modelo de inteligência artificial *Eliza*, desenvolvido por Joseph Weizenbaum, na década de 1960, que tinha por objetivo simular a figura de um psicoterapeuta rogeriano² em uma conversação entre pessoas com um *chatbot*, que acabou ganhando repercussão além do esperado diante da identificação das pessoas envolvidas com a máquina;
- b) investigar a dissonância cognitiva e a tendência humana à projeção antropomórfica no que tange a objetos inanimados;
- c) examinar as implicações da utilização de *chatbots* para fins de cuidados com a saúde mental ao direito à integridade psíquica, um direito fundamental e da personalidade, crucial para a preservação da dignidade humana.

A pesquisa utilizou o método hipotético-dedutivo, que consiste na escolha de proposições hipotéticas, e com certa viabilidade, para a tentativa de responder a um problema ou lacuna no conhecimento científico, com a realização de falseamento dessas hipóteses, para o fim de comprová-las ou não (Gil, 2008). A técnica de pesquisa utilizada foi a de revisão bibliográfica, a partir de obras, artigos de periódicos e legislação aplicável ao tema.

Como resultado, constatou-se que há uma tendência humana de antropomorfizar dispositivos tecnológicos, lendo respostas como resultados de emoções humanas, o que pode provocar apego, identificação e dependência, diante da percepção de reciprocidade, cenário que pode gerar ofensa à integridade psíquica. Há implicações éticas e jurídicas no que tange a sistemas de conversação para fins terapêuticos, sobretudo em relação ao direito à integridade psíquica. Destaca-se também, neste cenário, questões ligadas à privacidade das informações compartilhadas, à segurança e a vieses discriminatórios.

2 CHATBOTSE A INTERAÇÃO HUMANO-MÁQUINA

A utilização da Inteligência Artificial (IA) revolucionou diversas áreas do conhecimento e propiciou serviços e soluções baseadas no tratamento de dados pessoais

² Carl Rogers foi o criador do método terapêutico denominado “abordagem centrada na pessoa”, que se baseava na livre expressão e na autonomia do paciente: “é ele quem sabe o que lhe machuca e que direção tomar” (Nunes, 2013, p. 76-77).

e em informações abrangentes sobre o usuário, especialmente na área da saúde e dos cuidados com a saúde mental.

Os *chatbots* (*chatterbots*, *bots* ou agentes conversativos) são *softwares*³ criados para a conversação com seres humanos, de modo que podem ser utilizados como assistentes pessoais, para vendas, consultas ou resolução de diversos problemas (Oliveira, 2021; Ferrara *et al.*, 2016), por meio do Processamento de Linguagem Natural (PLN) (Abdul-Kader, 2015; Silva *et al.*, 2022). O uso de PLN envolve a manipulação computacional, a partir de palavras e diferentes estilos de escrita, um processo que intenta “entender” expressões humanas completas, para que o sistema seja capaz de devolver respostas úteis.

Portanto, refere-se ao processamento computacional automático de linguagem humana, o que inclui a utilização de algoritmos que usam o texto produzido pelo ser humano como entrada e outros que produzem texto com aparência natural como saída. Os *chatbots* atuais costumam utilizar a linguagem de programação *Python* e técnicas neurais (Goldberg, 2017; Oliveira, 2021, p. 28; Bird; Klein; Loper, 2009).

É possível utilizar *chatbots* para fins de *marketing*, como assistentes virtuais; para a customização em massa e de *Big Data*, no âmbito dos *influencers*; na área da Educação, os *chatbots* podem ser utilizados na assistência ao desenvolvimento de pessoas com espectro autista; na área da Saúde, podem ser utilizados pela telemedicina e para cuidados com pacientes.

O termo “*chatbot*” teria sido utilizado pela primeira vez em 1994, por Mauldin (1994), a partir da junção das palavras “*chat*” (do inglês *Conversational Hypertext Access Technology*⁴), em referência ao diálogo e à conversa entre duas ou mais pessoas, e “*bot*”, que é a abreviação de “*robot*” (“robô” em português) (Oliveira, 2021). Por meio de técnicas de processamento, como a NLTK para *Python*, podem ser aplicados para analisar a fala, bem como respostas inteligentes e apropriadas podem ser encontradas por meio da projeção de mecanismos (Abdul-Kader, 2015).

Um dos grandes desafios dos *chatbots* seria a “humanização” da conversa, em resposta às reações humanas sob os pontos de vista cognitivo e emocional. Diante disso, técnicas e ferramentas de IA foram desenvolvidas, como o *IBM Watson*, que é uma plataforma de construção de *bots*, a qual auxilia na aplicação de algoritmos cognitivos para a elaboração de conversas que seriam “mais humanas” e devolveriam respostas claras. Nesse contexto, destacam-se as plataformas *Chatfuel*, *Dialogflow* e *Bot Framework* (Gratzer; Goldbloom, 2020; Oliveira, 2021; Weizenbaum, 1966).

³ Como pontuam Stair e Reynolds (2011), os *softwares* podem ser conceituados como programas que comandam a operação do computador e permitem comandar o funcionamento de um sistema, com base em tarefas específicas.

⁴ Tradução livre: “Tecnologia de acesso a hipertexto conversacional”.

As interfaces de inteligência artificial, como o *Bing Chat*⁵, da *Microsoft*; o *ChatGPT*⁶, da *OpenAI*; e o *Bard*⁷, da *Google*, são fáceis de utilizar e parecem pessoais, já que, ao conversar com a IA, o usuário pode ter a sensação de que está conversando com outro ser humano. Conforme Picalho:

[...] é interessante pensar que tanto o *ChatGPT*, o *Bing Chat* e o *Bard* melhoram essa interação humano computador (IHC) popularizada pelas assistentes pessoais virtuais como *Siri*, *Alexa* e *Google Assistente* em formato de voz e tornam o diálogo muito convincente em termos de naturalidade. Mesmo não sendo necessariamente motores de busca, se assemelha muito a ideia popularizada pelo buscador *Ask Jeeves* nos anos 90 que vendia no imaginário do mecanismo de busca, a figura de um mordomo que responderia a todas as suas solicitações de pesquisa. Tudo aqui é uma troca. Ao mesmo tempo em que as pessoas testam essas ferramentas e/ou as incorporam no seu ambiente de trabalho, as empresas coletam esses feedbacks e aprimoram o modelo (Picalho, 2023, p. 50).

Alguns indivíduos relatam que desenvolveram sentimentos de apego em relação a determinados *chatbots*. “Outros ficaram tão convencidos da inteligência da IA que publicaram com confiança seus temores de sciência” (Sponheim, 2023a). Não é novo o fascínio por dispositivos mecânicos e de IA que imitam os seres humanos:

[...] os antigos egípcios construíam estátuas de divindades em madeira e pedra e as consultavam para obter conselhos. Os primeiros estudiosos budistas descreveram “pessoas de metal precioso” que recitavam textos sagrados. Os gregos antigos contavam histórias de Hefesto, o deus da ferraria, e seu amor pela criação de robôs. Ainda assim, foi somente nas décadas de 1940 e 50 que os computadores modernos começaram a trazer essas fantasias para mais perto do reino da realidade. À medida que os computadores se tornaram mais

⁵ O *Bing Chat* possui dados atuais e está atrelado aos resultados de pesquisa do buscador da *Microsoft*, tornando-o preciso e eficaz quanto a respostas referentes a acontecimentos recentes. “O *Bing Chat* permite que o usuário escolha um estilo de conversa dentre três disponíveis: mais criativo (com respostas mais originais e criativas); mais balanceado (com respostas mais informativas e amigáveis) e mais preciso (com respostas concisas e diretas) (Picalho, 2023, p. 44).

⁶ “No atual panorama, o *ChatGPT* configura-se como uma das mais recentes e sofisticadas ferramentas de Inteligência Artificial (IA) desenvolvidas pela empresa *OpenAI*, demonstrando capacidade para executar variadas tarefas com eficiência. O termo “*ChatGPT*” designa um modelo de linguagem fundamentado em IA que é empregado para gerar textos e realizar tarefas de processamento de linguagem natural”. A sigla “*GPT*” se refere aos “*Generative Pre-trained Transformer*” (em tradução literal: “Transformadores generativos pré-treinados), os quais descrevem a “arquitetura e a técnica de treinamento empregadas para a criação desses modelos de linguagem”, sendo modelos de redes neurais que geram textos semelhantes aos de humanos (Barreira; Moura-Silva; Gonçalves, 2024, p. 4).

⁷ O *Bard* é um *chatbot*, lançado pela empresa *Google*, que foi treinado por um conjunto gigante de dados textuais e códigos, mediante comunicação semelhante a uma conversa entre seres humanos, podendo ser descrita como uma Inteligência Artificial (IA) conversacional. Assim como o *Bing Chat*, o *Bard* está atrelado ao buscador da empresa e processa informações em tempo real por meio de pesquisa no *Google*. “O *Bard* é um grande modelo de linguagem do *Google AI* [...]. Ele pode gerar texto, traduzir idiomas, escrever diferentes tipos de conteúdo criativo e responder às suas perguntas de forma informativa. Ele ainda está em desenvolvimento, mas aprendeu a realizar muitos tipos de tarefas” (Picalho, 2023, p. 47).

potentes e difundidos, as pessoas começaram a ver o potencial das máquinas inteligentes (Sponheim, 2023a).

Em 1950, Alan Turing, matemático britânico, escreveu o seminário *Computing Machinery and Intelligence*⁸ e propôs o “teste de Turing”, por meio do qual uma pessoa conversava com um humano e com um robô localizados em salas diferentes, de modo que precisaria descobrir quem seria quem. Caso o robô fosse convincente, ele passaria no teste. Na visão de Turing, com o tempo, as pessoas fariam com máquinas inteligentes de forma consistente, sem esperar que fossem contrariadas (Sponheim, 2023a; Hall, 2019). À época, o teste levantou questões importantes sobre a mente, a consciência e a humanidade.

A partir da década de 1960, os cientistas da computação conseguiram desenvolver sistemas para planejar, racionalizar e perceber dados e informações. Foram criados programas que jogavam damas, resolviam problemas e provavam teoremas. Herbert Simon, pesquisador sobre IA, na década de 1980, previu que, um dia, as máquinas seriam capazes de fazer qualquer trabalho que uma pessoa pudesse fazer (Sponheim, 2023a).

Os primeiros *chatbots*, criados nas décadas de 1950 e 1960, eram muito limitados. Eles podiam “conversar sobre um tópico muito específico, como beisebol. Ao limitar o mundo de possíveis perguntas e respostas, os pesquisadores podiam criar máquinas que se passavam por ‘inteligentes’”, contudo, conversar com um sistema era como falar “com a Wikipédia”⁹, não com uma pessoa real” (Sponheim, 2023a).

O modelo de inteligência artificial *Eliza*¹⁰ foi desenvolvido por Joseph Weizenbaum, na década de 1960, para assumir o papel de uma terapeuta, com base em texto¹¹ (Bostrom, 2018; Oliveira, 2021). O professor do *Massachusetts Institute of Technology* (MIT) reconheceu que existiam padrões de fala na função de um terapeuta, logo, poderiam ser facilmente automatizados. A prática seguiria os princípios de Carl Rogers, criador do método terapêutico denominado “abordagem centrada na pessoa”, que se baseava na livre expressão e na autonomia do paciente: “é ele quem sabe o que lhe machuca e que direção tomar”¹² (Nunes, 2013, p. 76-77). O modelo de IA *Eliza*

⁸ Tradução livre: “Máquinas de Computação e Inteligência”.

⁹ A Wikipédia é uma “enciclopédia digital, colaborativa, de livre acesso e de atualização constante, que se caracteriza por compilar e administrar informações e saberes dispersos” (Varella; Bonaldo, 2020, p. 149).

¹⁰ Conforme Nunes (2013, p. 76), “o nome do programa foi inspirado na personagem Eliza Doolittle, da peça *Pigmaleão*, de George Bernard Shaw. A peça, que é inspirada no mito de *Pigmaleão*, por sua vez, inspirou também o filme *My Fair Lady* (EUA, 1964), um dos mais populares musicais norte-americanos do século XX”.

¹¹ O programa *Eliza* demonstrou que um computador poderia imitar um psicoterapeuta rogeriano (Bostrom, 2018).

¹² Para Rogers, “havia três condições básicas facilitadoras para que a abordagem tivesse resultados: a aceitação do outro de maneira positiva e incondicional, a empatia e a coerência do terapeuta” (Nunes, 2013, p. 77).

questionava o usuário: “*Alguma coisa está incomodando você?*”¹³. Diante da resposta, a inteligência identificava uma palavra-chave (*estou me sentindo triste*¹⁴) e a repetia em uma pergunta: “*É importante que você esteja se sentindo triste?*”¹⁵ ou “*Por que você está se sentindo triste?*”. Já quando o sistema não conseguia identificar uma simples palavra-chave, respondia com uma frase genérica, tal como: “*Por favor, continue*”, “*conte-me mais*”¹⁶ ou “*qual é a conexão disso, na sua opinião?*”¹⁷ (Sponheim, 2023a; Hall, 2019).

A partir da comunicação na função de psicoterapeuta, *Eliza* definia o contexto da conversa, sobretudo as expectativas dos usuários. O cenário de conversação, aliado a uma lógica de programação, levou as pessoas a tratarem a IA como se humana fosse. Várias pessoas confidenciaram segredos profundos ao “psicoterapeuta virtual”. A própria secretária de Joseph Weizenbaum pediu para ter conversas particulares com *ELIZA* (Nunes, 2013; Weizenbaum, 1966), que parecia ouvir e entender problemas humanos: “[...] os resultados, no entanto, o perturbaram”. As pessoas “pareciam ter conversas significativas com algo que ele nunca pretendia que fosse uma ferramenta terapêutica de fato. Para outros, porém, isso parecia abrir um mundo inteiro de possibilidades” (Sponheim, 2023a).

O fato de os *chatbots* parecerem humanos não significa que possam ser considerados bons produtos de IA (Sponheim, 2023a). O modelo de IA *Eliza* não entendia nada. Com base em uma conversa semelhante à humana, os indivíduos estavam predispostos a atribuir as próprias palavras e seus sentimentos ao programa. Como um espelhamento, *Eliza* simplesmente refletia os pensamentos de volta para as pessoas, isto é, a fascinação advinha, em parte, da utilidade do programa de retornar sentimentos, reformulando e reenquadrando opiniões.

Diante do potencial da IA *Eliza*, o pesquisador começou a promover a ideia de que o programa realmente poderia ser útil sob a perspectiva terapêutica, chamando a atenção da comunidade médica, já que o sistema poderia contribuir para expandir o acesso à saúde mental. As vantagens é que seria barato e que as pessoas poderiam de fato falar livremente com o *chatbot* (Sponheim, 2023a, 2023b).

Embora a ideia não tenha sido concretizada, outros pesquisadores que trabalhavam com saúde mental realizaram experimentos para investigar como poderiam utilizar o computador em seu trabalho. Destaca-se a criação, em 1972, do *chatbot* *Parry*, pelo psiquiatra Kenneth Colby, que simulava o estilo de conversação de uma pessoa com

¹³ No original: “*Is something troubling you?*”.

¹⁴ No original: “*I’m feeling sad*”.

¹⁵ No original: “*Is it important that you’re feeling sad?*” ou “*Why are you feeling sad?*”.

¹⁶ No original: “*Tell me more*”.

¹⁷ No original: “*Please go on*” ou “*What is the connection, you suppose?*”.

esquizofrenia paranoide e que, posteriormente, desenvolveu o programa integrativo “*Overcoming Depression*”¹⁸ (Sponheim, 2023).

Com o passar dos anos, Weizenbaum¹⁹⁻²⁰ se afastou de projetos expansivos da área, considerando que os campos que exigissem compaixão e compreensão humanas não poderiam ser automatizados. O autor se preocupava com a redução de algo tão íntimo, sutil e humano, como é o processo terapêutico, a códigos e à linguagem de programação, bem como com a possibilidade de que no futuro os *chatbots* enganassem de forma regular as pessoas, já que poderiam se apresentar como humanos em conversas. Ademais, previa a possibilidade de utilização da IA de forma distorcida por corporações e/ou governos. Nas décadas de 1980 e 1990, houve avanços quanto ao processamento de linguagem natural. Os cientistas passaram a utilizar métodos estatísticos, coletando documentos e padrões linguísticos. Já nas décadas de 2000 e 2010, os computadores foram treinados com a utilização das chamadas “redes neurais profundas”²¹, a partir da utilização de enormes quantidades de dados, que só se tornaram acessíveis com o surgimento da Internet.

Atualmente, os sistemas aprendem muito mais rápido e, mesmo que não consigam entender conceitos da forma como os humanos, podem ser considerados cada vez mais confiáveis, convincentes e pessoais (Hall, 2019). O *chatbot Woebot*²² foi desenvolvido pela *Woebot Labs* para a atuação na área de cuidados com a saúde mental de jovens estudantes universitários, tendo sido implementado por um grupo de pesquisa da Universidade de Stanford, nos Estados Unidos. O objetivo do sistema seria monitorar o humor, de modo

¹⁸ Tradução livre: “Superando a depressão”.

¹⁹ “Weizenbaum, ciente da condição rudimentar do robô *Eliza*, chocou-se com a repercussão de sua criação a ponto de repensar seus princípios. O programa não possuía o poder de compreensão a ele creditado por grande parte dos seus interlocutores. Aliás, ele não fora pensado como um potencial substituto de um psicoterapeuta humano – seu propósito era imitar somente” (Nunes, 2013, p. 77).

²⁰ Ele temia que os usuários não entendessem completamente que estavam conversando com vários circuitos e ponderou sobre as implicações mais amplas de máquinas que poderiam efetivamente imitar o senso de compreensão humana (Hall, 2019).

²¹ Do inglês “*deep neural networks*”, as redes neurais profundas são estruturas que são capazes de realizar uma simulação do funcionamento do sistema nervoso e do cérebro de seres vivos, sendo geralmente construídas de pequenos processadores, denominados “neurônios”. As redes neurais podem ser utilizadas na área da educação, para estimativas do mercado, a previsão de fluxo de indivíduos em redes de computadores, previsões climáticas etc. (Oliveira; Verson; Marcelino, 2018; Eremia; Liu; Edris, 2016; Matzavela; Chrysafiadi; Alepis, 2017; Luca; Gallo, 2017; Marzouq *et al.*, 2017; Vui *et al.*, 2013; Agrawal; Haidri, 2016). No cenário de criação de *chatbots*, as redes neurais são utilizadas para o desenvolvimento e o processamento de informações de entrada, já que se destacam por contribuírem para a apresentação de respostas mais precisas aos usuários durante a conversação. “Uma rede neural implementa um sistema de processamento composto de um grande número de pequenos elementos, denominados neurônios, que são interconectados de forma a resolverem problemas computacionais por meio de um processamento paralelo” (Oliveira; Veson, Marcelino, 2018; Eremia; Liu; Edris, 2016).

²² O *chatbot Woebot* está disponível para os sistemas Android e IOS por meio do link: WOEBOT: The Mental Health Ally. Google Play, 2024. Disponível em: <https://play.google.com/store/apps/details?id=com.woebot&hl=en&pli=1>. Acesso em: 5 jul. 2024.

que o indivíduo pudesse aprender sobre si mesmo, conduzindo exercícios baseados na Terapia Cognitivo-Comportamental²³, que auxiliam indivíduos a interromperem e reformularem padrões de pensamentos negativos, contribuindo para a diminuição de sintomas de depressão. O *chatbot* é transparentemente não humano, sendo interpretado por um avatar robô, e parte da sua personalidade é a curiosidade sobre os sentimentos humanos.

O *chatbot* pergunta às pessoas como elas estão se sentindo e o que está acontecendo com a vida delas no momento, na forma de conversação diária. O sistema conversa sobre saúde mental e bem-estar, encaminhando vídeos e outras ferramentas úteis, conforme o humor e as necessidades atuais dos usuários (Hall, 2019; Oliveira, 2021). Para Oliveira (2021), o sistema seria uma analogia a um livro de autoajuda, capaz de aprender sobre a pessoa e se tornar mais específico às suas necessidades a curto e longo prazo, tudo diante do processamento de linguagem natural, em conjunto com a técnica TCC, mediante interações que começam com um levantamento acerca do contexto do usuário (com, por exemplo, as seguintes perguntas: “O que está acontecendo no seu mundo agora?” ou “Como você está se sentindo?”).

Após a coleta de dados quanto ao humor do indivíduo, o sistema apresenta os conceitos básicos relacionados à Terapia Cognitivo-Comportamental, por meio de um *link* com vídeo ou breves “jogos de palavras”, projetados para facilitar o ensino sobre distorções cognitivas. Há uma introdução ao *bot*, que adverte que este, apesar de útil, não é totalmente capaz de entender as necessidades do usuário e que, embora haja monitoramento do sistema por psicólogos, isso não ocorre em tempo real, de modo que o serviço não deve ser utilizado como um substituto de um terapeuta. Em casos mais graves, o *chatbot* incentiva os participantes a acionarem os serviços emergência (ex.: 911 – número de emergência nos Estados Unidos). Após a realização de uma pesquisa com estudantes da Universidade de Stanford, verificou-se que os estudantes que utilizaram o

²³“A Terapia Cognitiva-Comportamental (TCC) é uma abordagem que leva em consideração diversos aspectos do paciente. De modo geral, essa psicoterapia busca compreender a forma como os indivíduos interpretam o seu arredor e como reagem a isso [...]. A Terapia Cognitiva-Comportamental (TCC) foi criada na década de 1960, pelo psiquiatra Aaron Beck, a partir das insatisfações que ele possuía com as teorias da época sobre a depressão. Beck começou a reparar que seus pacientes com depressão apresentavam pensamentos negativos e distorcidos sobre si mesmos, sobre o mundo e sobre os outros [...]. A ideia central da TCC é que ao mudar as crenças básicas do paciente sobre si mesmo, seu mundo e as outras pessoas; é possível alterar o modo como ele se sente, seus humores e seus comportamentos [...]. Isso significa que a Terapia Cognitiva-Comportamental tem como base o modelo cognitivo. Ou seja, acredita que a emoção e o comportamento são influenciados pela forma como o indivíduo interpreta os acontecimentos [...]. A TCC se fundamenta na crença de que o afeto e o comportamento de um indivíduo são em grande parte determinados pelo modo como ele estrutura o mundo em sua mente. As cognições do sujeito se baseiam em atitudes ou pressuposições desenvolvidas a partir de vulnerabilidades biológicas e experiências que ele teve ao longo de sua vida [...] as técnicas terapêuticas dessa abordagem visam identificar e testar reações distorcidas dos pacientes, guiando-os para a construção de comportamentos mais funcionais a sua realidade (Souza; Borges; Oliva, 2023).”

chatbot apresentaram melhora maior dos sintomas de depressão do que o grupo de estudantes que simplesmente leu um livro de autoajuda (Oliveira; 2021; Fitzpatrick; Darcy; Vierhile, 2017; Hall, 2019).

À medida que os usuários foram interagindo com o *chatbot*, novamente, seus desenvolvedores perceberam o *Efeito Eliza*: as pessoas estavam criando ligações emocionais com o programa, que envia *GIFs*²⁴, oferece cumprimentos e age mais como um incentivador do que como um terapeuta formal. Nos primeiros cinco dias, o sistema já havia sido utilizado por 50 mil usuários e trocava milhões de mensagens com pessoas todas as semanas. Allison Darcy, fundadora da *Woebot Labs*, acrescenta um ponto sobre o *Efeito Eliza*: há a possibilidade de que os humanos se envolvam em uma espécie de brincadeira quando interagem com os *chatbots*, isto é, não são necessariamente enganados, apenas ficam fascinados com o fato de que podem se ver refletidos em máquinas inteligentes (Hall, 2019).

O *ElizzBot*, que foi inicialmente denominado “Tess”, também é um *chatbot* de saúde mental e foi proposto para ser utilizado em momentos em que o usuário está tendo um ataque de pânico, precisa desabafar ou conversar antes de dormir. O usuário pode se conectar por meio de um aplicativo de mensagens instantâneas, como o *Facebook Messenger* ou por SMS (mensagem de texto pelo celular), com o *chatbot*, que responde imediatamente. As conversas guiadas pelo sistema abrangem questões ligadas ao bem-estar emocional, ao gerenciamento do estresse, à importância do autocuidado e da construção de resiliência diante de circunstâncias de luto e perda, de forma que podem servir também como suporte personalizado para profissionais e cuidadores familiares (Joerin; Rauws; Ackerman, 2019; Oliveira, 2021).

Já o *chatbot Owlle*²⁵, disponível no idioma francês, foi criado com o objetivo de disponibilizar ao público novas ferramentas de terapia. O *bot* utiliza experiências práticas de Psicologia, com terapias da área da Medicina baseadas em evidências (EBM), que sejam cientificamente avaliadas, utilizando um algoritmo de decisão e tecnologia fundamentada em regras (com correspondência de palavras). O *Owlle* realiza a escuta empática, propõe ferramentas/exercícios e promove a educação terapêutica. Pode ser utilizado para o gerenciamento de uma crise emocional, contribuindo para os seus usuários se responsabilizem por sua recuperação e tomem decisões que possibilitem melhora na qualidade de vida (Oliveira, 2021; Falala-Séchet *et al.*, 2019).

²⁴Os *GIFs* se tornaram culturalmente relevantes no contexto digital, sobretudo nas mídias sociais, sendo ferramentas que utilizam apelo sensorial por meio do uso de movimento, cor e repetição, uma vez que são considerados recursos adequados para gerar emocionalidade, pois podem representar uma série de sentimentos (Ash, 2015; Bourlai; Herring, 2014; Rúa-Hidalgo *et al.*, 2021).

²⁵A página do *chatbot Owlle* está disponível na rede social *Facebook*, por meio do link: OWLIE. Facebook, 2024. Disponível em: <https://www.facebook.com/owllechatbot>. Acesso em: 6 jul. 2024.

O *Vicent* é um *chatbot* que foi utilizado para pesquisas que envolvem o estudo da compaixão humana. Durante duas semanas, os participantes que conversavam com o sistema tinham que cuidar dele em um processo de autocompaixão. Esperava-se que, ao cuidar de *bot*, o usuário estimulasse a sua compaixão. De acordo com os idealizadores do experimento, a compaixão e a motivação são emoções-chave, que merecem ser melhor exploradas por meio de computação positiva e tecnologia voltada para o bem-estar (Oliveira, 2021; Lee *et al.*, 2019), uma vez que a:

[...] autocompaixão pode ajudar o bem-estar geral das pessoas por meio da bondade para com si mesmo, conexão com a humanidade e maior atenção plena. Apesar de um *chatbot* não ser a solução para curar problemas psicológicos e não se destinar à substituição da ajuda profissional, foi evidenciado que cuidar de um *chatbot* pode ajudar as pessoas a obterem maior autocompaixão do que quando são cuidadas por um *chatbot*. De acordo com as análises quantitativas e qualitativas, os autores sugerem que a interação humano-*chatbot* é uma área promissora para a Informática (Oliveira, 2021, p. 30-31).

O *chatbot Rumi* foi desenvolvido em 2017, a partir de diálogos baseados na Terapia Cognitiva Comportamental Focada em Ruminação²⁶ e em um modelo de alfabetização emocional, tendo por foco a conversação com adolescentes para fins de psicoeducação e melhoria dos sentimentos de depressão, ansiedade, ruminação e ideação suicida. A primeira versão do robô foi desenvolvida por meio da plataforma *Chatfuel* e funcionava por meio do *Messenger* do *Facebook* (Oliveira *et al.*, 2021; Oliveira, 2021).

O aplicativo de *chatbot Wysa* auxilia seus usuários a desenvolverem autoexpressão positiva, utilizando inteligência artificial empática para possibilitar um ambiente de autorreflexão externo e responsivo. O aplicativo é gratuito e disponível 24x7. O *app* responde às emoções expressadas pelo usuário por escrito e em suas conversas, utilizando práticas de autoajuda baseadas em evidências, como a TCC, a terapia comportamental dialética, a entrevista motivacional, o suporte de comportamento positivo, a atenção plena, o reforço comportamental e a resiliência emocional. O aplicativo possui um conselho consultivo que aprova o conteúdo e as ferramentas utilizadas. Os principais tópicos trabalhados são: ansiedade, foco, energia, relaxamento, sono, perdas, conflitos, preocupações etc. (Pandey; Shama; Wazir, 2022; Inkster; Sarda; Subramanian, 2018).

Apesar dos avanços da IA, pesquisadores e cientistas, como Melanie Mitchell, pontuam que um dos maiores desafios dos sistemas atuais é dominar a linguagem humana (“linguagem natural”), que, de certo modo, é quase equivalente ao sentimento. A

²⁶Sob a perspectiva da Terapia Cognitiva Comportamental (TCC), Watkins (2016) constatou que “o principal alvo primário para a diminuir e/ou prevenir os pensamentos e ações depressivas é focar na ruminação. O autor define ruminação como um pensamento recorrente e repetitivo sobre sintomas (fadiga, baixo humor), sentimentos, problemas, eventos perturbadores e aspectos negativos pessoais, tipicamente com foco nas causas, significados e implicações que rodeiam esses sentimentos” (Oliveira, 2021, p. 15).

linguagem reúne o conhecimento²⁷ sobre como o mundo funciona, o que engloba a compreensão de outros seres humanos e de conceitos fundamentais (Sponheim, 2023b).

Parte do intuito da IA hoje é desenvolver soluções úteis, significativas e inovadoras. Logo, não basta criar produtos que interajam como se fossem humanos, sendo necessário que os desenvolvedores compreendam o comportamento dos usuários (ex.: o motivo pelo qual agem dessa forma, como se divertem e o que faria a experiência ser mais prazerosa) (Sponheim, 2023a), de forma que a interação possa ser mais natural e intuitiva (Nunes, 2013).

Como pontua Sponheim (2023a), é comum que as pessoas rapidamente atribuam características semelhantes às dos humanos aos sistemas de inteligência artificial, o que reflete sua própria personalidade. As empresas claramente não veem “problema em fazer com que os *bots* pareçam humanos”, um exemplo disso é que a Meta “anunciou um conjunto completo de personalidades de IA com as quais os usuários podem interagir e aprender”.

Como pontua Sponheim (2023a) acerca da experiência do usuário (UX):

[...] melhorar as interfaces para que sejam agradáveis e fáceis de usar é exatamente o que os profissionais de UX vêm fazendo há décadas. Muitos dos conceitos fundamentais de UX e abordagens de pesquisa que foram estabelecidos pelo campo de UX se aplicam diretamente ao desenvolvimento de IA. O conforto e a leniência podem beneficiar substancialmente a experiência do usuário dos produtos. No entanto, ao realizar pesquisas para produtos de IA, os profissionais de UX devem estabelecer objetivos claros para o produto. O objetivo é estimular um sentimento de conexão emocional ou compreensão, ou otimizar a aprendizagem e a produtividade de um trabalhador? Os pesquisadores de experiência do usuário têm as habilidades necessárias para avaliar objetivamente os benefícios e as desvantagens de diferentes abordagens de interface de IA e podem fazer isso mantendo o foco na experiência dos usuários (Sponheim, 2023a).

Verifica-se que a base (e o desafio) da interação humano-*chatbot* é fundamentada na linguagem, na reciprocidade e na percepção de empatia, de forma que, com frequência, o ser humano compreende a IA a partir de valores humanos, com clara projeção antropomórfica, que será abordada na próxima seção.

3 DISSONÂNCIA COGNITIVA E PROJEÇÃO ANTROPOMÓRFICA

O experimento com o modelo de IA *Eliza*, conduzido por Weizenbaum na década de 1960, demonstra claramente a tendência humana de antropomorfizar os objetos com

²⁷“todo *chatbot* possui um interpretador e uma base de conhecimento – esta última seria como o cérebro do robô. O primeiro é responsável por intermediar o diálogo entre interator e o cérebro do robô – onde está a rede de associações do programa” (Nunes, 2013, p. 83).

os quais se relaciona, o que inclui animais²⁸, plantas e, no caso do presente artigo, os dispositivos tecnológicos, lendo comportamentos e respostas como resultados de uma “emoção humana”, ainda que estes sejam apenas resultados aleatórios, repetitivos ou mesmo vazios de sentido”. As condições para esse fenômeno são favorecidas “quando interlocutores desconhecem os princípios de programação ou a natureza automatizada das respostas” (Cleland, 2004; Nunes, 2013).

Weizenbaum denominou “*Efeito Eliza*” a projeção que as pessoas fizeram de que havia um ser humano por trás das respostas exibidas pela tela do sistema, desenvolvendo afeto e certa dependência emocional, um claro exemplo de resposta à dissonância cognitiva ocasionada pela relação homem-máquina, que é um viés que faz que as pessoas “procurem seguir suas crenças e ideologias, embora a realidade dos fatos a desminta”. Exemplo comum nos últimos anos é a propagação de pensamentos extremistas, negacionistas e de *fake news* no cenário midiático e virtual (Bicharra; Salgado; Pinto, 2023).

Nunes (2013, p. 79) destaca que, em uma conversa com um *chatbot*, o ser humano tende a preencher de sentido qualquer resposta elaborada pelo sistema. “Desconhecendo o poder do programa, diante de uma resposta tão subjetiva quanto uma mudança de olhares”, o interator, que é uma espécie de espectador-participante e que interage de forma tanto ativa quanto fisicamente, dá-se por entendido.

Como observa Balkin (2015), a projeção de emoções humanas em objetos inanimados não é um fenômeno que pode ser considerado recente na história da humanidade, contudo, quando tal situação envolve a utilização de robôs, pode acarretar várias consequências, entre elas:

- a) quanto maior a proteção antropomórfica, mais as pessoas atribuem culpa ao robô, e não à pessoa que o utiliza;
- b) a presença de robôs no contexto de sistemas de vigilância amplia a sensação subjetiva de que o indivíduo está sendo observado;
- c) os seres humanos tendem a assumir riscos para preservar a integridade de robôs alvo da projeção do que ela relação a coisas que podem ser consideradas meras ferramentas;
- d) os humanos podem sofrer danos emocionais distintos diante da perda de seus companheiros robóticos.

Quanto ao robô humanoide *Sophia*, que recebeu cidadania na Arábia Saudita, Negri (2020) observa que, apesar dos traços antropomórficos, este segue sendo um sistema que pode ser considerado simples:

²⁸Quanto aos animais, Cavalcanti Neto e Pimentel (2024, p. 140-141) consideram que as pessoas são propensas ao antropomorfismo, ou seja, o ser humano projeta características “a outras entidades para fazê-las parecer mais humanas. Esse efeito aumenta quando os animais exibem um comportamento que associamos mais prontamente com cognição e emoções”.

[...] a metáfora pode se transmutar em uma falácia: a aparência humana pode nos levar a pensar em robôs como pessoas. Assim, como nem todos os robôs são andróides, a ilusão provocada pelo antropomorfismo da forma pode ser perigosa quando pensamos em iniciativas regulatórias pautadas em falsas suposições sobre a capacidade dos próprios artefatos robóticos. A projeção de características humanas em robôs não depende da sua forma. Mesmo quando um artefato robótico não tem formato antropomórfico, as pessoas projetam nessas tecnologias qualidades humanas, como consciência e inteligência. À medida que a autonomia do sistema aumenta, dificultando conexões entre as entradas (seus comandos) e o comportamento do robô, as analogias com os seres humanos são reforçadas, o que, por sua vez, pode prejudicar qualquer tentativa normativa, seja no que se refere ao debate da ética, seja em questões jurídicas, como a determinação de quem seria o responsável por possíveis danos causados pelos artefatos robóticos (Negri, 2020, p. 4).

Quanto à IA, os usuários sabem que estão interagindo com uma máquina, equipada com um *software*. Contudo, cada vez mais, os sistemas respondem com a “credibilidade que só um humano poderia ter”, palco propício para uma dissonância cognitiva, diante da eventualidade de atribuição de humanidade ao *chatbot*, o que, no âmbito da Psicologia, é conhecido como projeção antropomórfica, já que cria uma ilusão na mente humana de que há presença de empatia (Hall, 2019). O indivíduo que interage com a IA passa a utilizar o sistema como se estivesse lidando com um amigo próximo, com reciprocidade emocional, o que pode criar um senso de familiaridade e gerar apego e identificação. “O *chatbot* estaria associado ao que é projetado nele, modificando-se a percepção que se tem dele a ponto de esquecer que é um programa de *software*” (Bicharra; Salgado; Pinto, 2023):

[...] a despeito das limitações que estão invariavelmente envolvidas na difícil tarefa de fazer compreender os acordos tácitos de uma conversa, nós – interlocutores – atribuímos uma suposta inteligência humana a estes sistemas, pois é naturalmente do nosso desejo que sejamos compreendidos de maneira recíproca (Nunes, 2013, p. 81).

Ao se deparar com essas máquinas, o ser humano parece querer acreditar que são pensantes e se preocupa em imbuir estas com qualidades humanas (Carr, 2011; Nunes, 2013). É estabelecido um jogo que, tal como ocorreu com o modelo *Eliza*, abre “espaço para a ambiguidade da real natureza dos interlocutores” (Nunes, 2013, p. 81). Bostrom (2018, p. 255), quanto à possibilidade de existência de uma superinteligência²⁹⁻³⁰ no

²⁹Destaca-se que, no âmbito da União Europeia, a Resolução do Parlamento Europeu de 16 de fevereiro de 2017 apresentou orientações sobre Robótica, com uma proposta de criação de uma “personalidade eletrônica” para artefatos robóticos “inteligentes”. Para a mencionada Resolução, um robô poderia ser considerado inteligente se possuisse as seguintes características: a) sensores capazes de permitir a troca de dados com o ambiente; b) capacidade de aprendizado a partir da experiência e da interação com o meio; c) suporte material; d) capacidade de adaptação; e) ausência de vida na acepção biológica. Recomendou-se a constituição de um registro próprio, com a formação de regimes de seguros e fundos de compensação, bem como a criação de um estatuto jurídico de robôs para artefatos complexos (Negri, 2020; União Europeia, 2017).

³⁰Pontua-se que há autores, como Tena (2023), que visualizam a perspectiva futura de existência de mecanismos de IA que possam evoluir para a condição de Ser, o que exigiria o reconhecimento como

futuro, adverte quanto ao perigo exatamente da projeção antropomórfica:

[...] é importante que não se antropomorfize a superinteligência quando pensamos em seus potenciais impactos. Tal perspectiva antropomórfica encorajaria o surgimento de expectativas infundadas a respeito da trajetória de crescimento de uma IA embrionária, além de questões acerca da psicologia, motivações e capacidades de uma superinteligência madura. Por exemplo, uma suposição muito comum é que uma máquina superinteligente seria semelhante a um ser humano muito inteligente, mas nerd. Nós temos a tendência de especular que uma IA teria uma inteligência técnica, mas lhe faltaria traquejo social, ou que ela seria extremamente lógica, embora pouco intuitiva e criativa. Essas ideias advêm, muito provavelmente, da observação: nós olhamos para os computadores atuais e constatamos que eles são muito bons de cálculo, possuem ótima memória e seguem regras ao pé da letra, porém são alheios ao contexto e sutilezas sociais, normas, emoções e política. Tal associação é reforçada quando observamos que pessoas que trabalham bem com computadores tendem a ser, elas mesmas, “nerds”. É, então, natural assumir que uma inteligência computacional mais avançada venha a ter atributos semelhantes, mas em maior escala.

Bicharra, Salgado e Pinto (2023) mencionam que o *ChatGPT* (*Generative Pre-trained Transformer*)³¹ é atualmente o mais moderno dos *chatbots*, já que é um robô de conversação “desenvolvido pela empresa *OpenAI* que usa um modelo de geração de linguagem natural baseado em modelos pré-treinados que vão se ajustando usando a técnica de aprendizado não supervisionado”, ou seja, encontra “padrões de relações e conceitos nos textos disponíveis na grande base de textos não previamente classificados da *Web* para gerar respostas às perguntas feitas”. O autor pontua que, apesar de não gerar perguntas (fator essencial em um diálogo), a qualidade das respostas do *ChatGPT* é impressionante. Em estudo, Abreu descreve o que o **Chat respondeu** sobre os impactos psicológicos da interação humano-máquina, sobretudo para indivíduos com depressão (Bicharra; Salgado; Pinto, 2023):

[...] a interação humano-*chatbot* pode ter impactos psicológicos significativos, especialmente para pessoas com depressão. A depressão é uma doença mental que afeta a forma como as pessoas se sentem, pensam e se comportam, e pode levar a sentimentos de tristeza, perda de interesse, falta de energia e dificuldade em se relacionar com os outros. Os *chatbots* podem ser úteis para as pessoas com depressão, pois eles podem fornecer suporte emocional e orientação de forma anônima e acessível. Eles podem ser programados para fornecer respostas compreensivas, encorajadoras e orientadoras, e podem ser acessados 24 horas por dia, 7 dias por semana. Além disso, os *chatbots* podem ser usados como uma forma de ajudar as pessoas a se expressarem e a processar seus sentimentos. No entanto, é importante lembrar que os *chatbots* não são substitutos para o

Pessoa IA (Pessoa Inteligência Artificial). O surgimento de uma Pessoa IA movimentaria o sistema antropocêntrico ao ensinar direitos e deveres que poderiam ser atribuídos a este Ser tecnológico diante da apresentação de algum nível de senciência rudimentar.

³¹Tradução livre: “Transformador Generativo Pré-Treinado”.

tratamento médico ou psicológico. Eles não podem diagnosticar ou tratar doenças mentais, e as pessoas com depressão devem ser encaminhadas para profissionais de saúde qualificados para o tratamento. Além disso, os *chatbots* podem ser limitados na sua capacidade de entender e lidar com questões complexas ou nuances emocionais. Em geral, a interação humano-*chatbot* pode ser uma ferramenta útil para ajudar as pessoas com depressão a se sentirem menos sozinhas e mais compreendidas. No entanto, é importante lembrar que eles não podem substituir o cuidado médico e psicológico e deve ser usado como complemento ao tratamento (Bicharra; Salgado; Pinto, 2023).

Quanto às questões sociais, éticas e jurídicas que precisam ser consideradas para que os *chatbots* sejam utilizados de maneira responsável, o **ChatGPT** destacou os seguintes pontos:

Privacidade: os usuários podem compartilhar informações pessoais com os *chatbots*, e há preocupações com a segurança dessas informações e como elas são coletadas, armazenadas e utilizadas.

Transparência: os usuários podem não ser conscientes de que estão falando com um *chatbot* e não com uma pessoa real, e isso pode levar a mal-entendidos ou a confiança desnecessária.

Responsabilidade: se um *chatbot* comete um erro ou causa danos, pode ser difícil determinar a quem atribuir a responsabilidade.

Impacto no emprego: os *chatbots* podem ser usados para automatizar tarefas que antes eram feitas por seres humanos, o que pode levar ao desemprego.

Bias: os *chatbots* podem replicar os preconceitos e desigualdades presentes no conjunto de dados usado para treiná-los, o que pode levar a respostas imprecisas ou discriminatórias (Bicharra; Salgado; Pinto, 2023).

Shijie, Yuxiang e Qinghua (2023) destacam que, do estágio inicial do *chatbot Eliza* até o recente *ChatGPT*, vários aplicativos demonstraram avanço significativo quanto à compreensão, à geração e ao raciocínio no que tange à linguagem humana. A experiência natural de interação humano-IA fornece aos usuários serviços de informação convenientes, bem como levanta questões quanto à credibilidade, especialmente a partir de uma análise sociotécnica, com dimensões de explicabilidade, geração de conteúdo, interfaces de terceiros, visão antropomorfista de considerar a IA como um ator social³², e da consideração quanto a cuidados com a alfabetização dos usuários.

Shah *et al.* (2016) realizaram experimento comparando cinco sistemas de diálogo artificial, quais sejam, *Cleverbot*, *Elbot*, *Eugene*, *Goostman*, *IFred* e *Ultra Hal*, com a versão

³²“Os chamados robôs sociais são pensados exatamente para interagirem com seres humanos em ambientes não controlados. Para tanto, intensificaram-se os estudos e projetos que buscam o desenvolvimento de artefatos capazes de interagir com as pessoas de forma mais natural possível. Os robôs sociais se caracterizam pela possibilidade, ainda que aparente, de transmitir emoções, incentivar e formar relacionamentos sociais, demonstrar personalidade, usar pistas naturais de comunicação e interagir socialmente com as pessoas. Já existe um campo próprio de estudo denominado interação humano-robô (IHR), que procura, a partir da valência social, replicar nos artefatos robóticos uma variedade de sinais e marcadores presentes na comunicação humana, como expressões faciais e até mesmo a linguagem (Negri, 2020, p. 6-7).

on-line de *Eliza*. Os resultados apontaram que os sistemas de conversação modernos utilizam técnicas para iniciar e manter o aprendizado do diálogo com base em interações com humanos pela Internet. Os sistemas analisados representam uma melhoria do modelo *Eliza*, já que, incorporados à rede, proporcionam interação vinte e quatro horas por dia. Hodiernamente, os sistemas aprendem com a maneira como os humanos conversam, sendo capazes de compartilhar opiniões pessoais, transmitir experiências, conversar sobre dramas familiares, serem relevantes, ao mesmo tempo que vagos e, eventualmente, enganarem, assim como os seres humanos.

Estima-se que a indústria de aplicação dos *chatbots* atinja US\$200 bilhões até 2026 (Brown, 2022; Han; Yin; Zhang, 2023). Além dos avanços em IA, como o processamento e a compreensão de linguagem e natural, esse crescimento é explicado pelo aumento da adesão das pessoas às plataformas conversacionais, que foi impulsionado pela difusão de aplicativos disponíveis, como a *Siri*³³, da *Apple*, e o *Bixby*³⁴, da *Samsung* (Bicharra; Salgado; Pinto, 2023; Nobles *et al.*, 2020; Santos; Polianov, 2021).

Uma das problemáticas quanto ao uso dos *chatbots* é a sua aplicação para o fim de representação de pessoas mortas, por meio de avatares digitais, de modo que os familiares e os amigos poderiam interagir com o ente querido já falecido. Como destacam Bicharra, Salgado e Pinto (2023):

[...] se pensarmos na dificuldade das pessoas de lidarem com as perdas de familiares e amigos, utilizar técnicas de IA pode ser uma boa ideia para encaminhar a aceitação do que Sigmund Freud chamou de transitoriedade das coisas e pessoas. Para Freud, numa visão otimista, ter consciência da efemeridade da vida, das coisas e pessoas nos ajuda a elaborar as perdas e ver que a “exigência de imortalidade, por ser tão obviamente um produto dos nossos desejos, não pode reivindicar seu direito à realidade; o que é penoso pode, não obstante, ser verdadeiro.” Já numa visão pessimista, recorrendo a uma alegoria com um poeta, para Freud a finitude das coisas e pessoas acabaria por tirar delas seu valor e beleza [...]. Aqueles que defendem o uso de *chatbots* de pessoas mortas defendem que esse diálogo post-mortem ajudaria a realizar a finitude e o processo de luto se consubstanciaria nessa interação. Logo, procuram caminhar na direção otimista da noção de transitoriedade freudiana. Entretanto, quanto mais o tipo de *chatbot* sugere a possibilidade de um relacionamento profundo e pessoal, mais o padrão instintivo de apego emocional parece ser solicitado e, quanto mais a reciprocidade, mais o antropomorfismo intencional do *chatbot* tem a chance de ser eficaz [...].

³³A *Siri* é uma assistente pessoal digital da empresa *Apple*, sendo uma espécie de artefato comunicacional que auxilia usuários a executarem tarefas do dia a dia de maneira automatizada, mediante linguagem natural falada e escrita. Trata-se de robô humanizado, como personagem feminina (Santos; Polianov, 2021).

³⁴O *Bixby* é um assistente pessoal virtual presente em dispositivos da marca *Samsung* e possui diversas funções, entre elas, a emissão de comandos de voz, o reconhecimento de objetos, os lembretes e o controle de outros devices inteligentes (Nobles *et al.*, 2020).

Além de todos os dilemas éticos envolvidos na utilização de dados de pessoas mortas para simular conversas, “o risco de dependência emocional ou alienação deve, portanto, ser considerado, sob pena de se ter consequências psíquicas e distúrbios relacionados à maneira de se lidar com as perdas e o luto” (Bicharra; Salgado; Pinto, 2023).

O conhecimento acerca do impacto dos *chatbots* nos contextos individual, grupal e social é limitado. A pesquisa de Croes *et al.* (2022) indicou que, diante do fluxo de conversas com os *chatbots*, os humanos, impulsionados pelas perguntas feitas pelos sistemas, tendem a iniciar um processo de autorrevelação (*self-disclosure*), já que passam informações pessoais e íntimas aos robôs. Entretanto, “os *chatbots* não correspondem e não se autorrevelam, o que pode dificultar a formação de um relacionamento humano-*chatbot*”³⁵ (Bicharra; Salgado; Pinto, 2023).

O debate no que tange à possibilidade de utilização da IA para fins terapêuticos ganhou força na última década. É necessário pontuar o desafio ético, de responsabilidade e cultural quanto ao uso de *chatbots*, já que, se tais tecnologias “imitam o comportamento humano e são criadas por pessoas com diferentes visões do mundo: os agentes conversacionais podem ser tendenciosos e prejudiciais, assim como as pessoas podem” (Bicharra; Salgado; Pinto, 2023). Constata-se, portanto, que o futuro desta tecnologia e seu potencial impacto exigem pesquisas interdisciplinares, conectadas à realidade sociopolítica, cultural e de saúde das partes envolvidas.

4 A UTILIZAÇÃO DE *CHATBOTS* NO ÂMBITO DOS CUIDADOS COM A SAÚDE MENTAL: IMPLICAÇÕES AO DIREITO DA PERSONALIDADE À INTEGRIDADE PSÍQUICA

Importa a este trabalho a análise das implicações do *Efeito Eliza* e da projeção antropomórfica ao direito à integridade psíquica, sobretudo diante de eventual cenário de utilização de *chatbots* para fins terapêuticos e de saúde mental.

Inicialmente, pontua-se que, no Brasil, a saúde consta nos arts. 6º e 196 e seguintes da Constituição Federal de 1988 como um direito social³⁶ e passou a ser considerada direito de todos e dever do Estado, já que este assumiu a obrigação de assegurá-la por meio

³⁵“Os bate-papos da rede Internet – os *chats* – são formas sociais bastante definidas e consolidadas em rede: contatos rápidos, curtos e truncados. Fala-se de tudo, mas especialmente da própria convivência, visto que o encobrimento típico destas conversas nos dá a privacidade que precisamos para compartilhar sentimentos e impressões pessoais” (Nunes, 2013, p. 87).

³⁶“Os direitos sociais conferem a noção de ser fundamental para proteger o indivíduo, já segundo acontecia no pensamento clássico dos direitos de liberdade, mas agora com a salvaguarda da instituição, um fato social muito mais rico e receptivo à participação criativa e à valorização da personalidade que o pano de fundo clássico da solidão individual, no qual nascera a exaltação do homem não concreto e isolado, sem a intensidade dos valores existenciais, os únicos efetivados de maneira plena pelo social” (Otero; Massarutti, 2016, p. 852).

de políticas públicas sociais e econômicas que visem à redução do risco de doenças e outros agravos. O acesso à saúde no país é de caráter universal e igualitário quanto às ações e aos serviços para a sua promoção, proteção e recuperação (Brasil, 1988).

Sob ponto de vista do Direito Internacional, instrumentos internacionais, como a Declaração Universal de Direitos Humanos³⁷, proclamam o direito humano a um padrão de vida suficiente para uma vida digna, o que inclui os direitos à saúde, ao bem-estar e a cuidados médicos (ONU, 1948). No mesmo sentido, o Pacto Internacional sobre os Direitos Econômicos, Sociais e Culturais³⁸ dispõe que toda pessoa tem o direito de “desfrutar o mais elevado nível possível de saúde física e mental” (OEA, 1966). A saúde, por ser um direito de âmbito coletivo, faz parte dos direitos considerados de segunda geração/dimensão, que englobam os direitos sociais, culturais e econômicos, além dos direitos coletivos, inseridos no contexto do Estado Social, logo, são direitos que necessitam da intervenção do Estado para impor limites ao abuso econômico, à exploração das pessoas e do trabalho, incluindo a prestação de benefícios (Bonavides, 2014; Otero; Massarutti, 2016).

Otero e Massarutti (2016) e Szaniawski (2002) destacam que a saúde é um direito fundamental que possui relação com o pleno desenvolvimento da personalidade humana e é crucial para uma perspectiva mínima da vida digna, nos termos do princípio da dignidade da pessoa humana (art. 1º, inc. III, da CF/88) (Brasil, 1988).

Segundo a definição da Organização Mundial da Saúde (OMS), a saúde mental é um estado de bem-estar mental que permite que as pessoas “lidem com o estresse da vida, percebam suas habilidades, aprendam e trabalhem bem e contribuam para sua comunidade”, assim, seria um direito humano básico, crucial para o desenvolvimento pessoal, comunitário e socioeconômico. A saúde mental é mais do que a ausência de transtornos mentais, sendo uma experiência complexa e vivenciada de forma diferente por cada pessoa, com graus de dificuldade, sofrimento, resultados sociais e clínicos potencialmente diversos (WHO, 2022a, tradução nossa).

O bem-estar de uma pessoa não depende somente do aspecto psicológico e emocional, sendo influenciado também por fatores, como a saúde física, o apoio social e as condições de vida. Portanto, a saúde mental é também determinada por aspectos ambientais, econômicos e sociais. A garantia do direito constitucional à saúde inclui o cuidado com a saúde mental, sendo um dever do Estado oferecer condições dignas de

³⁷Declaração Universal dos Direitos Humanos. Artigo 25. 1. “Todo ser humano tem direito a um padrão de vida capaz de assegurar a si e à sua família **saúde, bem-estar**, inclusive alimentação, vestuário, habitação, **cuidados médicos** e os serviços sociais indispensáveis e direito à segurança em caso de desemprego, doença invalidez, viuvez, velhice ou outros casos de perda dos meios de subsistência em circunstâncias fora de seu controle” (ONU, 1948, grifo nosso).

³⁸Pacto Internacional sobre os Direitos Econômicos, Sociais e Culturais. Artigo 12 §1. “Os Estados-partes no presente Pacto reconhecem o direito de toda pessoa de desfrutar o mais elevado nível de saúde física e **mental**” (OEA, 1966, grifo nosso).

cuidado em saúde³⁹ aos cidadãos (Brasil, 2024). A integridade física e a integridade psíquica, bem como a dignidade fazem parte da personalidade do ser, de forma que o desrespeito ao direito à saúde atinge esses bens fundamentais, uma vez que a vida deve ser gozada com um mínimo de qualidade. Diante disso, um cenário de ausência de recursos e de condições precárias de saúde implica sério risco de extinção da própria vida.

Nesse sentido, é a higidez psíquica que propicia ao ser humano a capacidade de se desenvolver e exercer seus direitos de modo livre e consciente, bem como a possibilidade de comandar suas escolhas, seus desejos e seu próprio projeto de vida. Pontua-se que é um dos pressupostos da dignidade humana a “afirmação da integridade física e espiritual da pessoa humana como dimensão irrenunciável de sua individualidade autonomamente responsável” (Canotilho, 2003, p. 363). “Os atributos psíquicos do ser humano estão relacionados aos sentimentos de cada indivíduo. A própria noção de saúde passa pela higidez mental”. Assim, uma vez violada a integridade psíquica, “sepulta-se a perspectiva da vida do ser humano (e não apenas sua existência), juntamente com a sua dignidade”⁴⁰. Dessa forma, o dever de não violação à integridade psíquica opõe-se tanto individualmente quanto em relação à comunidade, restando a obrigação de não interferência no “aspecto interno da personalidade de outrem” (Castro; Ramiro; Tamaoki, 2023, p. 16540-16541):

[...] a psique humana e o seu uso tratam-se de um direito e uma faculdade que são atribuídos ao ser humano pela própria natureza, mas que são reconhecidos pelo ordenamento jurídico positivo que, por sua vez, dá o permissivo para o uso da referida faculdade e ainda determina a sua proteção, classificado no ordenamento positivo brasileiro, entre os Direitos da Personalidade, como direito à integridade psíquica (Castro; Ramiro; Tamaoki, 2023, p. 16553).

No ordenamento jurídico brasileiro, a integridade psíquica é compreendida como um elemento crucial e íntimo da personalidade, voltada para o se interior. A integridade psíquica é que possibilita ao indivíduo o equilíbrio emocional para o desenvolvimento da personalidade ao longo da vida (Castro; Ramiro; Tamaoki, 2023; Rodotà, 2014; Macedo; Rocha, 2020).

³⁹No Brasil, a política de saúde mental é pautada em princípios como o cuidado, a desinstitucionalização e os direitos humanos. As pessoas em situação de crise podem ser atendidas por meio da Rede de Atenção Psicossocial (RAPS), formada por vários serviços, com características e finalidades distintas, no âmbito do Sistema Único de Saúde (SUS), regido pelos princípios de universalidade, integralidade e equidade, já que o tendo em vista o dever de prestar atendimento acessível, amplo e justo a todos (Brasil, 1988, 2024).

⁴⁰“Trata-se, portanto, do direito que tem cada pessoa de manter incólume seus atributos psíquicos, sua higidez mental e consequentemente, a sua dignidade, onde qualquer conduta atentatória à tal integridade considera-se ilícita [...]. São ilícitas, portanto, as condutas que violam e afetam a integridade psíquica, que causam sentimentos negativos e desagradáveis, como tristeza, vergonha, constrangimento etc.” (Castro; Ramiro; Tamaoki, 2023, p. 16540).

Para Bittar (1999), a integridade psíquica é um direito da personalidade assegurado com o escopo de preservar a incolumidade da psiquê, o conjunto psicoafetivo e pensante da estrutura humana. Quanto aos direitos da personalidade, Tepedino (2004) pontua que estes são essenciais à tutela da pessoa humana, considerando a proteção da sua dignidade e da integridade. Os direitos da personalidade, portanto, tutelam o que o ser humano possui de mais ímpar, isto é, suas características individuais, que não subsistem diante de imposições arbitrárias ou desrespeito aos limites de interferência na esfera individual, bem como diante de ofensa à sua integridade, seja física, seja mental.

Conforme Bittar (1999, p. 64), os direitos da personalidade se qualificam a partir de “caracteres bem definidos, tratando-se de direitos inatos, absolutos, extrapatrimoniais, intransmissíveis, imprescritíveis, impenhoráveis, vitalícios, necessários e oponíveis *erga omnes*”. São os direitos da pessoa considerada em si mesma e anteriores ao Estado. Como bem observa de Cupis (1961, p. 17), existem certos direitos sem os quais a personalidade “restaria uma susceptibilidade completamente irrealizada, privada de todo o valor concreto: direitos sem os quais todos os outros direitos subjetivos perderiam todo o interesse para o indivíduo”. Assim, caso não existissem, o indivíduo não existiria como tal.

No Brasil, os direitos da personalidade são tratados em capítulo próprio pelo Código Civil de 2002 (Capítulo II), entre os arts. 11 e 21. Segundo tal diploma legal, os direitos da personalidade são intransferíveis e irrenunciáveis, não podendo o seu exercício sofrer limitação voluntária (art. 11, CC/02) (Brasil, 2002; Tobbin; Cardin, 2021). Szaniawski (2002), Moraes (2008) e Tepedino (2004) compreendem que o rol de direitos da personalidade disposto no Código Civil não é taxativo, de forma que outros direitos, não contemplados pelo *codex*, também são fundamentais para o desenvolvimento da personalidade humana, sobretudo diante da evolução social e da dificuldade de o Direito acompanhar e regular todas as esferas e as temáticas da ordem social ao tempo que estas são identificadas e reconhecidas.

No Brasil, alguns autores asseveram que a dignidade da pessoa humana, prevista no art. 1º, inciso III, da Constituição Federal, anunciada como um dos fundamentos da República Federativa do Brasil, seria a cláusula geral de proteção da personalidade, protegendo o ser em sua totalidade em face de situações que implicassem ofensa a sua individualidade, cuja tutela é essencial para o desenvolvimento da personalidade (Szaniawski, 2002; Tobbin; Cardin, 2021).

Groeninga (2006) pontua que o direito à integridade psíquica poderia ser considerado o mais importante dos direitos da personalidade, tendo em vista que é o psiquismo que permite a qualidade humana:

[...] a própria noção de saúde passa pela higidez mental. A ideia de dignidade humana carrega em si um desejado equilíbrio psicológico. São ilícitas, portanto, as condutas que violam e afetam a integridade psíquica, que causam

sentimentos negativos e desagradáveis, como tristeza, vergonha, constrangimento etc. (Bessa; Reis, 2020, p. 13).

Como analisam Ramiro, Fachin e Tamaoki (2022, p. 51-52), trata-se do direito que tem cada pessoa de manter incólume seus atributos psíquicos, sua higidez mental e, conseqüentemente, a sua dignidade”, logo, “qualquer conduta atentatória à tal integridade considera-se ilícita”. Stoco (2007) afirma que o dano psíquico ocorre diante de distúrbio ou perturbação causada à pessoa, por meio de sensações desagradáveis, mesmo que passageiras.

Quanto à evolução tecnológica e sua faceta neoliberal, os autores destacam que, “levando o homem a se autoexplorar em razão da relação de docilidade imposta também sobre as mentes, torna-se evidente como as novas técnicas de poder apresentadas pela psicopolítica neoliberal violam frontalmente à integridade psíquica”. Ramiro, Fachin e Tamaoki (2022, p. 51-52) apontam que:

[...] o desafio apresentado pela evolução das técnicas de poder reside na compreensão de que a complexa relação existente entre a nova anatomia política e a integridade psíquica do homem se dá a partir de meandros profundos e muito sutis que necessitam ser explorados e pesquisados a fim de trilhar caminhos para se alcançar a proteção integral da pessoa humana na contemporaneidade⁴¹ [...] a anatomia política encontra sua base não só na aplicação de disciplinas sobre os corpos, mas também sobre as mentes. O adestramento almejado, hoje, atinge a dimensão psíquica do ser humano. A realidade atual supera a antiga: a liberdade e a positividade é que mantêm o sujeito disciplinado, perseguindo a máxima produtividade como um empreendedor de si mesmo. Formam-se, portanto, não apenas corpos dóceis, mas mentes dóceis (Ramiro; Fachin; Tamaoki, 2022, p. 52-53).

Castro, Ramiro e Tamaoki (2023, p. 16535) analisam os desafios da proteção legal do direito à integridade psíquica no Brasil, diante de eventual deficiência na conceituação deste pela doutrina brasileira, pelo Poder Judiciário⁴² e quanto ao tratamento legislativo

⁴¹Para Zigmunt Bauman, “não existiria uma “identidade verdadeira”, já que somos incessantemente forçados a torcer e moldar nossas identidades ou adotarmos novas identidades, sem ser permitido que nos fixemos a uma delas, mesmo querendo. É importante contextualizar que a noção de identidade na contemporaneidade trazida por Bauman está bastante próxima da ideia de persona – assumir um personagem de um dado contexto. Para Bauman, tanto as relações humanas quanto nossas identidades, por extensão, estão sob uma lógica de consumo – assim como produtos que são adquiridos e descartados” (Nunes, 2013).

⁴²Os autores analisaram o tratamento dado ao direito à integridade psíquica no Brasil pelo Supremo Tribunal Federal (STF) e pelo Superior Tribunal de Justiça (STJ). Após a análise de 29 acórdãos, observaram que os referidos tribunais reconhecem a importância da proteção do direito à integridade psíquica, contudo, não tecem maiores esclarecimentos sobre a sua conceituação e seus contornos, logo, não foi encontrada abordagem aprofundada sobre o assunto (decisões analisadas pelo STF: “RHC 55240; HC 67836; RHC 140754 AgR; RHC 128726; ADI 5995; HC 95142; HC 89238; Ext 1434; HC 91386; Inq 3932; ADI 2404; HC 123734; HC 123533; e ADPF 54; decisões analisadas no âmbito do STJ: REsp n. 1.884.887/DF; HC n. 661.801/SP; REsp n. 1.820.891/PR; RHC n. 108.350/RN; REsp n. 1.543.465/RS; HC n. 442.456/SP; REsp n. 1.637.884/SC; REsp n. 1.636.815/DF; REsp n.

adequado às mudanças sociais contemporâneas. Em relação aos direitos da personalidade, os autores pontuam a necessidade de especial estudo quanto à integridade psíquica, sobretudo diante dos “reflexos da pós-modernidade, apesar de recaírem nos mais diversos campos sociais, ocasionam severos reflexos na psique humana”.

Quanto à utilização de *chatbots* para fins de saúde mental, verifica-se que, tendo em vista que grande parcela da população mundial ainda não possui condições de acesso a tratamentos e a cuidados nesse contexto, estes poderiam significar certa democratização de informações e exercícios baseados em técnicas psicoterapêuticas, sobretudo para fins de controle de sintomas como estresse, ansiedade, depressão e ideação suicida.

Os *chatbots*, muitas vezes, servem de porta de entrada para a terapia convencional, evidenciando a necessidade de cuidados e indicando que os usuários busquem um terapeuta humano. Muito embora a maioria dos *apps* de saúde mental sejam abandonados poucos dias após a instalação pelo usuário, há evidências que sugerem que os *bots* podem aumentar o engajamento e o prazer com os cuidados com a saúde mental no âmbito digital (Coghan *et al.*, 2023; Prochaska *et al.*, 2023; Silveira; Paravidini, 2024).

Como pontuam Lucas *et al.* (2014), as pessoas tendem a divulgar mais informações sensíveis durante entrevistas clínicas realizadas por computador do que quando estas são realizadas por seres humanos⁴³. Os modelos recentes de IA podem, de forma inovadora, ampliar as interações terapêuticas e estabelecerem relações de empatia e confiança, simulando uma espécie de “aliança terapêutica” com os seus usuários (Silveira; Paravidini, 2024). Assim, seria importante que tais sistemas não se apresentassem com um terapeuta, um avatar ou qualquer figura que tivesse representação humana e pudesse ampliar a projeção antropomórfica, já que o contrário disso seria o mesmo que simular algo, que é uma forma de provocar engano (Grodziewicz; Hohol, 2023). Destaca-se que, na Bélgica, em março de 2023, um homem cometeu suicídio após semanas de conversação intensa com o *chatbot* *Eliza* (Walker, 2023).

O desenvolvimento e a manutenção de modelos como os *chatbots* devem seguir princípios como o da não maleficência, da beneficência, da autonomia, da justiça e da explicabilidade (que envolveria transparência e responsabilidade). Para isso, é crucial avaliar os riscos e os benefícios da utilização dos *chatbots*, estudar e divulgar evidências de sustentem sua eficácia e segurança, respeitando a privacidade e o consentimento dos usuários, para garantir a qualidade e a acessibilidade dos serviços, o que também exige

1.593.857/MG; AgRg no AREsp n. 395.426/DF; EDcl no AREsp n. 656.318/MG; REsp n. 1.349.385/PR; CC n. 126.175/PE; AgRg no REsp n. 1.159.867/MG; REsp n. 392.712/PR” (Castro; Ramiro; Tamaoki, 2023, p. 16541).

⁴³Um teste realizado com 97 participantes comparou as respostas dos indivíduos diante da interação com outro ser humano e com agentes virtuais. Os usuários reportaram que eram capazes de relevarem mais informações sobre o que estavam sentindo e sobre seus sentimentos, inclusive a tristeza, a agentes virtuais (Lucas *et al.*, 2014; Silveira; Paravidini, 2024).

monitoramento e avaliação de resultados. É necessário considerar a revelação de crimes às autoridades, o acionamento de serviços de emergência diante de risco iminente aos usuários ou terceiros e a fatores, como a diversidade e a particularidade das experiências dos indivíduos (Coghlan *et al.*, 2023; Silveira; Paravidini, 2024).

Rana e Singh (2023) abordam que uma das principais limitações de tais sistemas é a falta de dados suficientes sobre certos transtornos mentais, uma vez que as ferramentas de IA são dependentes de grandes quantidades de dados para que possam fornecer diagnósticos confiáveis, de modo que a falta de dados pode limitar sua eficácia. Há também falta de transparência e responsabilização sobre o uso de inteligência artificial para cuidados com a saúde mental. É crucial que os pacientes estejam cientes de como seus dados são coletados, armazenados e tratados, tendo a capacidade de controlar como estes podem ser utilizados (Rana; Singh, 2023). As informações coletadas pelos *bots* envolvem a privacidade dos usuários e seriam dados sensíveis⁴⁴. Tal cenário evidencia o problema já pontuado por autores como Zuboff (2021) acerca da opacidade de decisões tomadas por sistemas inteligentes e do tratamento de dados no contexto do Capitalismo de Vigilância, especialmente em termos de monetização e diante de eventual vazamento de dados. Há um risco de que as informações compartilhadas com tais sistemas sejam utilizadas ou compartilhadas sem o consentimento do usuário.

Sedlakova e Trachsel (2023) ressaltam a falta de consenso quanto a precauções éticas e a critérios para o desenvolvimento dos *chatbots*. De igual modo, não há uma normativa ou um órgão específico para analisar se tais sistemas estão sendo desenvolvidos de forma ética. Pontua-se que um modelo de IA não possui o mesmo nível de dever e responsabilidade de um terapeuta humano, já que a IA não é um “sujeito moral” e não há regulamentação legal clara quanto aos maus resultados ou eventuais dados deste tipo de relação estabelecida.

Possati (2023) evidencia que o desenvolvimento deve ser pautado na responsabilidade de agentes (usuários, desenvolvedores, *designers*, programadores, proprietários e *software*) e pacientes (que interagiriam com os sistemas). O autor pontua que não seria possível haver responsabilidade real sem uma relação entre o agente e o paciente, porque o primeiro poderia não agir de forma ética em relação à demanda do usuário.

Como observam Silveira e Paravidini (2024, p. 13-14):

[...] uma das questões que emerge é como se teoriza a responsabilidade do criador (desenvolvedor, proprietário, designer e software) das possíveis falhas e

⁴⁴Nos termos do art. 5º, inc. II, da Lei Geral de Proteção de Dados Pessoais (LGPD): “Para os fins desta Lei, considera-se: [...] II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural (Brasil, 2018).

danos causados pela sua criação (*chatbots* e IAs relacionais). As discussões sobre a responsabilidade dos pais para com as manifestações subjetivas dos seus filhos estão presentes nas discussões psicanalíticas e servem de base para ver como esse problema é mais sério do que é apresentado pelos pesquisadores citados neste trabalho. O que seria de responsabilidade, que tipo de organização se faz necessária para o desenvolvimento de inteligências artificiais cada vez mais capazes de simular uma consciência e sendo assim um artefato de desejo parece atravessar o sujeito de modo a lançá-lo numa instância de potência narcísica, tendo em vista a capacidade de ampliar os limites do próprio sujeito em sua relação com a realidade, bem como balançar a estrutura narcísica causando medo e desconforto, como se de alguma forma o avanço dessas tecnologias fosse evidenciando cada vez mais as incapacidades humanas. Nesse ponto, a construção com base ética de IAs deve se pautar em articulação com a ética psicanalítica, a fim de que novas formas de pensar tais artefatos possam existir, possibilitando um avanço no desenvolvimento de tecnologias articuladas nas relações sujeito e máquinas.

Os mencionados autores, retomando os estudos de Sigmund Freud, pontuam que não é papel do analista satisfazer a vontade do analisando. Já quanto a modelos de IA, questionam: “seria a IA capaz de encarnar a demanda ao Outro, satisfazendo assim para sustentar a obediência aos ideais sociais?”. “Qual seria o prejuízo do “acesso a uma ferramenta que está em total disponibilidade ao usuário atendendo a todas as suas demandas”? Os *chatbots* podem não só atender às demandas do usuário, oferecendo soluções para seus conteúdos psíquicos, bem como angústias e dores, mas também frustrar ou desafiar o indivíduo, levando-o a questionar ou modificar seus padrões de comportamento e pensamento (Silveira; Paravidini, 2024, p. 13).

Diante dessas considerações, verifica-se que a utilização de *chatbots* para fins de cuidados com a saúde mental deve ser encarada com seriedade e responsabilidade. A conversação com modelos de IA pode contribuir para a transmissão de informações relacionadas ao bem-estar físico e emocional, sobretudo porque as ferramentas estão disponíveis a qualquer tempo. Contudo, destaca-se que tais serviços não possuem o condão de substituir a figura de um terapeuta humano, sobretudo sob o ponto de vista da responsabilidade e atuação ética. Ressalta-se que tal utilização pode criar no usuário a ilusão de que há a presença de empatia na conversa, em razão de eventual projeção antropomórfica. O indivíduo pode interagir com o modelo de IA como se estivesse lidando com um amigo próximo, acreditando que há reciprocidade emocional, o que pode criar um senso de familiaridade e gerar apego e identificação, circunstância que ofenderia o seu direito da personalidade à integridade psíquica.

5 CONCLUSÃO

O trabalho teve por objetivo analisar as implicações do Efeito Eliza e da projeção antropomórfica ao direito da personalidade à integridade psíquica diante da eventual

utilização de *chatbots* para fins de cuidados com saúde mental.

A partir da pesquisa conduzida por Weizenbaum, na década de 1960, verifica-se que o escopo do modelo de inteligência artificial utilizado pelos *chatbot* tem por objetivo propiciar interação e conversação com os humanos de forma natural e intuitiva para a solução de uma gama de problemas que podem ser resolvidos com a observação de dados e padrões de comportamento humano.

Nos últimos anos, as interações humanas com sistemas de inteligência artificial (IA) cresceram de forma exponencial, sobretudo diante da democratização do acesso à Internet e da criação de modelos de *chatbots* cada vez mais ágeis e eficientes, como é o caso do ChatGPT. A utilização da IA revolucionou diversas áreas do conhecimento e propiciou serviços de informações abrangentes, sobretudo na área da saúde e dos cuidados com a saúde mental.

O fenômeno “Efeito Eliza” descreve a tendência humana a atribuir a sistemas inteligentes capacidades e sentimentos humanos, em especial diante da interação com *chatbots*, provocada por uma dissonância cognitiva, isto é, é comum que pessoas passem a tratar sistemas de IA como se pessoas fossem, diante de uma falsa percepção de reciprocidade e empatia, contexto que poderia gerar apego e dependência, o que traz implicações ao direito à integridade psíquica, um direito da personalidade essencial à proteção da dignidade humana.

Apesar de inovadores, eventuais sistemas de IA que sejam utilizados para fins de saúde mental podem ser imprecisos, vagos e limitados, tendo em vista a tendência humana de também atribuir sentido a todo que lhe é dito ou não dito. Destaca-se a falta de transparência quanto à privacidade e à segurança dos dados coletados, especialmente porque são de conteúdo pessoal e íntimo, e a possibilidade de vieses discriminatórios, uma vez que os sistemas de IA e os aplicativos atuais estão, em regra, estabelecidos sob a ótica e a conveniência do mercado tecnológico, com viés neoliberal. Logo, há possibilidade de manipulação e intermediação de interesses de terceiros, estranhos a uma conversação para fins terapêuticos.

O trabalho pontuou que há uma limitação em tais sistemas quanto a dados sobre certos transtornos mentais, fator que é essencial para que possam contribuir com diagnósticos confiáveis. É essencial que os usuários estejam cientes e possuam controle acerca de como os seus dados são coletados, armazenados e tratados, já que as informações coletadas pelos *bots* envolvem a privacidade dos usuários e seriam dados sensíveis. Contudo, destaca-se que há risco de que tais dados sejam utilizados ou compartilhados sem o consentimento do usuário, sobretudo em um contexto de monetização de dados pessoais.

O desenvolvimento e a manutenção desses modelos devem ser pautados em princípios éticos (entre eles: não maleficência, beneficência, autonomia, justiça e

explicabilidade - que envolveria transparência e responsabilidade por parte de seus agentes). A forma de identificação do *chatbot* também pode gerar repercussão: imitar um terapeuta é tão questionável quanto se apresentar como um avatar deslocado das figuras representativas de cuidados com saúde, cenário que pode gerar apego, identificação e dependência, tendo em vista, novamente, a tendência de projeção e espelhamento, o que comprometeria o livre desenvolvimento da personalidade do indivíduo, o seu direito à integridade psíquica e, conseqüentemente, a sua dignidade.

REFERÊNCIAS

ABDUL-KADER, Sameera A. Survey on chatbot design techniques in speech conversation systems. **International Journal of Advanced Computer Science and Applications**, v. 6, n. 7, p. 72-80, 2015.

AGRAWAL, Shubham; HAIDRI, Shruthi N. Murtaza. A neural conversational model for *chatbots*. **International Journal of Research in Computer and Communication Technology**, v. 5, n. 10, p. 516-519, 2016.

ASH, James. Sensation, Networks, and the GIF: Toward an Allotropic Account of Affect. In: HILLS, Ken; PAASONEN, Susanna; PETIT, Michael (ed.). **Networked Affect**. Boston: The MIT Press, 2015. p. 119-134.

BALKIN, Jack. The path of robotics law. **California Law Review Circuit**, Berkeley, v. 6, n. 1, p. 50-56, June 2015.

BARREIRA, Jonas Souza; MOURA-SILVA, Marcos Guilherme; GONÇALVES, Tadeu Oliver. Desafios e oportunidades para o uso do Chat GPT na pesquisa em educação matemática. **Revista Ibero-Americana de Estudos em Educação**, Araraquara, v. 19, n. esp. 2, p. 1-22, 2024.

BESSA, Leonardo Roscoe; REIS, Milla Pereira Primo. Dano moral e dor: direito autônomo à integridade psíquica. **Civilistica.com**, v. 9, n. 1, p. 1-17, 2020.

BICHARRA, Ana Cristina; SALGADO, Luciana; PINTO, Fernando. Tem alguém aí? Reflexões sobre a interação humano-*chatbot*. **Horizontes**, 1 fev. 2023. Disponível em: <https://horizontes.sbc.org.br/index.php/2023/02/tem-alguem-ai-reflexoes-sobre-a-interacao-humano-chatbot/>. Acesso em: 9 jul. 2024.

BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural language processing with Python**. Sebastopol: O'Reilly, 2009.

BITTAR, Eduardo Carlos Bianca. Direitos do consumidor e direitos da personalidade: limites, intersecções, relações. **Revista de Informação Legislativa**, Brasília, v. 36, n. 143, p. 63-70, jul./set. 1999.

Efeito Eliza: implicações da projeção antropomórfica à integridade psíquica diante da utilização de *chatbots* para cuidados com a saúde mental

BONAVIDES, Paulo. **Curso de Direito Constitucional**. 29. ed. São Paulo: Malheiros, 2014.

BOSTROM, Nick. **Superinteligência**: Caminhos, perigos, estratégias. São Paulo: Darkside Books, 2018.

BOURLAI, Elli; HERRING, Susan C. Multimodal communication on Tumblr: “I Have So Many Feels!”. In: ACM CONFERENCE ON WEB SCIENCE, Bloomington, 2014. **Anais [...]**. Nova Iorque: ACM, 2014. p. 171-175.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 fev. 2024.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Brasília, DF: Presidência da República, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/110406compilada.htm. Acesso em: 8 fev. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 5 jun. 2024.

BRASIL. Ministério da Saúde. **Saúde Mental**. 2024. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/saude-mental>. Acesso em: 6 jun. 2024.

BROWN, Elizabeth L. Bots with feelings: study explores deploying AI chatbots with emotions in customer service. **Muma College of Business**, 6 dez. 2022. Disponível em: <https://www.usf.edu/business/news/2022/12-6-chatbot-study-denny-yin.aspx>. Acesso em: 6 jun. 2024.

CANOTILHO, José Joaquim Gomes. **Direito Constitucional e Teoria da Constituição**. 7. ed. São Paulo: Almedina, 2003.

CARR, Nicholas. **A geração superficial**: o que a Internet está fazendo com o nosso cérebro. Rio de Janeiro: Agir, 2011.

CASTRO, Matheus Felipe de; RAMIRO, Marcus Geandré Nakano; TAMAOKI, Clara Carrociní. Desafios contemporâneos na proteção jurídica da integridade psíquica no Brasil. **Cuadernos de Educación y Desarrollo**, v. 15, n. 12, p. 16533-16561, 2023.

CAVALCANTI NETO, Gabriel de Oliveira; PIMENTEL, Alexandre Freire. A proteção dos robôs sociais em equiparação aos animais. **Revista Eletrônica Direito & TI**, Porto Alegre, v. 1, n. 18, p. 140-161, 2024.

CLELAND, Kathy. Talk to Me: getting personal with interactive art. In: EDMONDS, Ernest; GIBSON, Ross (ed.). **Interaction: systems, practice and theory**. Sydney: University of Technology, 2004.

CROES, Emmelyn A. J.; ANTHEUNIS, Marjolijn L.; GOUDBEEK, Martijn B.; WILDMAN, Nathan W. "I Am in Your Computer While We Talk to Each Other" a Content Analysis on the Use of Language-Based Strategies by Humans and a Social Chatbot in Initial Human-Chatbot Interactions. **International Journal of Human-Computer Interaction**, v. 39, n. 10, p. 2155-2173, 2022.

COGHLAN, Simon; LEINS, Kobi; SHELDERICK, Susie; CHEONG, Marc; GOODING, Piers; D'ALFONSO, Simon. To chat or bot to chat: Ethical issues with using chatbots in mental health. **Digital Health**, Londres, v. 9, n. 1 p. 1-10, June 2023.

CUPIS, Adriano de. **Os direitos da personalidade**. Tradução de Adriano Vera Jardim e Antonio Miguel Caeiro. Lisboa: [s. n.], 1961.

EREMIA, Mircea; LIU, Chen-Ching; EDRIS, Abdel-Aty. **Advanced Solutions in Power Systems: HVDC, FACTS and Artificial Intelligence**. Nova Jersey: Wiley-IEEE Press, 2016.

FALALA-SÉCHET, Clara; ANTOINE, Lee; THIRIEZ, Igor; BUNGNER, Catherine. *Owlle*: a chatbot that provides emotional support for coping with psychological difficulties. In: INTERNATIONAL CONFERENCE ON INTELLIGENT VIRTUAL AGENTS, 19., 2019, Paris. **Anais [...]**. Paris: IVA 19, 2019. p. 236-237. Disponível em: <https://dl.acm.org/doi/10.1145/3308532.3329416>. Acesso em: 6 jul. 2024.

FERRARA, Emilio; VAROL, Onur; DAVIS, Clayton; MENCZER, Filippo; FLAMMINI, Alessandro. **Communications of the ACM**, v. 59, n. 7, p. 96-104, 2016.

FITZPATRICK, Kathleen Kara; DARCY, Alison; VIERHILE, Molly. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. **JMIR, Mental Health**, v. 4, n. 2, 2017.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GOLDBERG, Yoav. Neural network methods for natural language processing. In: HIRST, Graeme (ed.). **Synthesis Lectures on Human Language Technologies**. Portico: Morgan & Claypool Publishers, 2017.

GRATZER, David; GOLDBLOOM, David. Therapy and e-therapy: preparing future psychiatrists in the era of apps and chatbots. **Academic Psychiatry**, p. 1-4, 2020.

GRODNIOWICZ, J. P.; HOHOL, Mateusz. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. **Frontiers in Psychiatry**, Laussane, v. 14, n. 1, May 2023.

GROENINGA, Giselle Câmara. O direito à integridade psíquica e o livre desenvolvimento da personalidade. In: PEREIRA, Rodrigo da Cunha (org.). **Família e Dignidade Humana**: V Congresso Brasileiro de Direito de Família. São Paulo: IOB Thompson. 2006. p. 439-455.

HALL, Delaney. Episode 382: The ELIZA Effect. **99% Invisible**, 12 out. 2019. Disponível em: <https://99percentinvisible.org/episode/the-eliza-effect/>. Acesso em: 9 maio 2024.

HAN, Elizabeth; YIN, Dezhi; ZHANG, Han. Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service? **Information Systems Research**, v. 34, n. 3, 2023.

INKSTER, Becky; SARDA, Shubhankar; SUBRAMANIAN, Vinod. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. **JMIR Mhealth Uhealth**, v. 6, n. 11, p. 1-14, 2018.

JOERIN, Angela; RAUWS, Michiel; ACKERMAN, Mary Lou. Psychological artificial intelligence service, tess: Delivering on-demand support to patients and their caregivers: technical report. **Cureus**, v. 11, n. 1, 2019.

LEE, Minha; ACKERMANS, Sander; AS, Nena van; CHANG, Hanwen; LUCAS, Enzo; IJSSELSTEIJN, Wijnand A. Caring for Vincent: a chatbot for self-compassion. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2019, Glasgow. **Anais [...]**. Glasgow: CHI, 2019.

LUCA, Giuseppina de; GALLO, Mariano. Artificial neural networks for forecasting user flows in transportation networks: Literature review, limits, potentialities and open challenges. In: INTERNATIONAL CONFERENCE ON MODELS AND TECHNOLOGIES FOR INTELLIGENT TRANSPORTATION SYSTEMS, 5., 2017. **Anais [...]**. IEEE, 2017. Disponível em: <https://ieeexplore.ieee.org/document/8005644/>. Acesso em: 9 jul. 2024.

LUCAS, Gale M.; GRATCH, Jonathan; KING, Aisha; MORENCY, Louis-Philippe. It's only a computer: virtual humans increase willingness to disclose. **Computers in Human Behavior**, Amsterdam, v. 37, n. 1, p. 94-100, Aug. 2014.

MACEDO, Larissa de Alencar Pinheiro; ROCHA, Maria Vital da. O plano de parentalidade como instrumento de salvaguarda da integridade psíquica da criança e do adolescente. **Duc In Altum**, Recife, v. 28, n. 12, p. 7-29, dez. 2020.

MAULDIN, Michael L. Chatterbots, Tinymuds, and the Turing Test Entering the Loebner Prize Competition. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 12., Seattle, 1994. **Anais [...]**. Seattle: AAAI, 1994.

MARZOUQ, Manal; FADILI, Hakim El; LAKHLIAI, Zakia; ZENKOUAR, Khalid. A review of solar radiation prediction using artificial neural networks. In:

INTERNATIONAL CONFERENCE ON WIRELESS TECHNOLOGIES, EMBEDDED AND INTELLIGENT SYSTEMS (WITS), 2017, Fez. **Anais [...]**. Fez: IEEE, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7934658>. Acesso em: 6 jun. 2024.

MATZAVELA, Vasiliki; CHRYSAFIADI, Konstantina; ALEPIS, Efthimios. Questionnaires and artificial neural networks: a literature review on modern techniques in education. In: GLOBAL ENGINEERING EDUCATION CONFERENCE, 2017, Atenas. **Anais [...]**. Atenas: EDUCON, 2017. p. 1700-1704. Disponível em: <https://ieeexplore.ieee.org/document/7943077>. Acesso em: 6 jul. 2024.

MORAES, Maria Celina Bodin de. Ampliando os direitos da personalidade. In: VIEIRA, José Ribas (org.). **20 anos da Constituição cidadã de 1988: efetivação ou impasse institucional?** Rio de Janeiro: Forense, 2008. p. 369-388.

NEGRI, Sergio Marcos Carvalho Avila. Robôs como Pessoas: a personalidade eletrônica na Robótica e na inteligência artificial. **Pensar: Revista de Ciências Jurídicas**, v. 25, n. 3, 2020.

NOBLES, Alicia L.; LEAS, Eric C.; CAPUTI, Theodore L.; ZHU, Shu-Hong; Strathdee, Steffanie A.; AYERS, John W. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. **NPJ Digital Medicine**, n. 3, 2020. Disponível em: <https://www.nature.com/articles/s41746-019-0215-9>. Acesso em: 6 jun. 2024.

NUNES, Fabio Oliveira. Conversação eletrônica em trabalhos de arte e tecnologia. **Texto Digital**, v. 9, p. 76-90, 2013.

OLIVEIRA, Alana Lucia Souza. **Um chatbot psicoeducativo para elucidar e diminuir sintomas de ansiedade, ruminação e depressão**. 2021. 75 f. Dissertação (Mestrado em Ciência da computação) – Universidade Federal de Sergipe, São Cristóvão, SE, 2021. Disponível em: https://ri.ufs.br/bitstream/riufs/14560/2/ALANA_LUCIA_SOUZA_OLIVEIRA.pdf. Acesso em: 6 jul. 2024.

OLIVEIRA, Alana Lucia Souza; MATOS, Leonardo Nogueira; COLAÇO JUNIOR, Methanias; DELABRIDA, Zenith Nara Costa; MURGANTE, Beniamino; BLEČIĆ, Ivan; MISRA, Sanjay; GERVASI, Osvaldo; TANIAR, David; ROCHA, Ana Maria A. C.; TARANTINO, Eufemia; GARAU, Chiara; APDUHAN, Bernady O.; TORRE, Carmelo Maria. An initial assessment of a chatbot for Rumination-Focused Cognitive Behavioral Therapy (RFCBT) in College Students. In: COMPUTATIONAL SCIENCE AND ITS APPLICATIONS, 2021. **Anais [...]**. ICCSA, 2021. p. 549-564. Disponível em: https://discovered.ed.ac.uk/discovery/fulldisplay?docid=cdi_springer_books_10_1007_978_3_030_86979_3_39&context=PC&vid=44UOE_INST:44UOE_VU2&lang=en&search_scope=UoE&adaptor=Primo%20Central&tab=Everything&query=null,vi,%2036%20pages%20;%2020%20cm. Acesso em: 9 fev. 2024.

OLIVEIRA, Graceline; VENSON, Ramon; MARCELINO, Roderval. Redes neurais aplicadas no desenvolvimento de *chatbots*: uma análise bibliométrica. **Artefactum: Revista de Estudos em Linguagem e Tecnologia**, v. 10, n. 2, 2018.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Assembleia Geral das Nações Unidas. **Declaração Universal dos Direitos Humanos**. 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 6 jun. 2024.

ORGANIZAÇÃO DOS ESTADOS AMERICANOS (OEA). **Pacto Internacional dos Direitos Econômicos, Sociais e Culturais**. Adotada pela Resolução n.2.200-A (XXI) da Assembleia Geral das Nações Unidas, em 16 de dezembro de 1966 e ratificada pelo Brasil em 24 de janeiro de 1992. 1966. Disponível em: <https://www.oas.org/dil/port/1966%20Pacto%20Internacional%20sobre%20os%20Direitos%20Econ%C3%B3micos,%20Sociais%20e%20Culturais.pdf>. Acesso em: 6 jun. 2024.

OTERO, Cleber Sanfelici; MASSARUTTI, Eduardo Augusto de Souza. Em conformidade com o direito fundamental à saúde previsto na Constituição brasileira de 1988, é possível exigir do Estado a prestação de fosfoetanolamina sintética para pessoas com câncer? **Revista Jurídica Cesumar – Mestrado**, v. 16, n. 3, 2016.

OWLIE. **Facebook**, 2024. Disponível em: <https://www.facebook.com/owlielechatbot>. Acesso em: 6 jul. 2024.

PANDEY, Sumit; SHARMA, Srishti; WAZIR, Samar. Mental healthcare chatbot based on natural language processing and deep learning approaches: Ted the therapist. **International Journal of Information Technology**, v. 14, p. 3757-3766, 2022.

PICALHO, Antonio Carlos. O que acontece quando um bibliotecário pergunta ao CHATGPT como ele deveria ser utilizado na biblioteca?: um teste com os *chatbots* GPT-3.5, Bing Chat e Bard. **Revista Bibliomar**, v. 22, n. 2, p. 39-51, 2023.

POSSATI, Luca M. **Unconscious networks: philosophy, psychoanalysis, and artificial intelligence**. Nova Iorque: Routledge, 2023.

PROCHASKA, Judith J.; VOGEL, Erin A.; CHIENG, Amy; BAIOCCHI, Michael; PAJARITO, Sarah; PIRNER, Maddison; DARCY, Alison; ROBINSON, Athena. A relational agent for treating substance use in adults: protocol for a randomized controlled trial with a psychoeducational comparator. **Contemporary Clinical Trials**, Amsterdam, v. 127, n. 1, p. 1, Apr. 2023.

RAMIRO, Marcus Geandré Nakano; FACHIN, Jéssica Amanda; TAMAOKI, Clara Carrocini. A docilidade das mentes e a ameaça à integridade psíquica. **Direito & Paz**, v. 16, n. 4, p. 41-55, 2022.

RANA, Usha; SINGH, R. The role of artificial intelligence in mental health care. **SocArXiv Papers**, 2023. Disponível em: <https://osf.io/preprints/socarxiv/r4umy>. Acesso em: 22 set. 2024.

RODOTÀ, Stefano. **El derecho a tener derechos**. Madri: Trotta, 2014.

RÚA-HIDALGO, Idoia; GALMES-CEREZO, Maria; CRISTOFOL-RODRÍGUEZ, Carmen; ALIAGAS, Irene. Understanding the Emotional Impact of GIFs on Instagram through Consumer Neuroscience. **Behavioral Science**, v. 11, n. 8, p. 1-15, 2021.

SANTOS, Luiza Carolina dos; POLIANOV, Beatriz. O que têm em comum Alexa, Siri, Lu e Bia? Assistentes digitais, sexismo e rupturas de performances de gênero. **Galáxia**, São Paulo, n. 46, p. 1-24, 2021.

SEDLAKOVA, Jana; TRACHSEL, Manuel. Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? **The American Journal of Bioethics**, Londres, v. 23, n. 5, p. 4-13, May 2023.

SHAH, Huma; WARWICK, Kevin; VALLVERDÚ, Jordi; WU, Defeng. Can machines talk? Comparison of Eliza with modern dialogue systems. **Computers in Human Behavior**, v. 58, p. 278-295, 2016.

SHIJIE, Song; YUXIANG, Zhao; QINHGUA, Zhu. From ELIZA to ChatGPT: AI-Generated Content (AIGC) Credibility Evaluation in Human-Intelligent Interactive Experience. **Information and Documentation Services**, v. 44, n. 4, p. 35-42, 2023.

SILVA, Carlos Henrique Nascimento Domingues da; DIAS, Douglas Vieira; SOARES, Guilherme Augusto Cardoso; ROMÃO, Ricardo Fusano; CAMPOS FILHO, Amadeu Sá. O uso dos *chatbots* como suporte ao aprendizado para urgência e emergência: uma revisão integrativa. **Revista de Saúde Digital e Tecnologias Educacionais**, Fortaleza, v. 7, n. 1, p. 1-13, 2022.

SILVEIRA, Paulo Victor dos Reis; PARAVIDINI, João Luiz Leitão. Ética da aplicação de inteligências artificiais e chatbots na saúde mental: uma perspectiva psicanalítica. **Revista Pesquisa Qualitativa**, v. 12, n. 30, p. 1-16, abr. 2024.

SOUZA, Renata de; BORGES, Karen; OLIVA, Daniel. O que é a Terapia Cognitiva-Comportamental (TCC) e como ela funciona? **Hospital Israelita Albert Einstein**, 6 jun. 2023. Disponível em: <https://vidasaudavel.einstein.br/o-que-e-a-terapia-cognitiva-comportamental-tcc-e-como-ela-funciona/>. Acesso em: 9 jul. 2024.

SPONHEIM, Caleb. O Efeito Eliza: por que amamos a IA. **Fernando Giannini**, 2023a. Disponível em: <https://fernandogiannini.com.br/o-efeito-eliza-por-que-amamos-a-ia/>. Acesso em: 9 jul. 2024.

SPONHEIM, Caleb. The ELIZA Effect: Why We Love AI. **NN Group**, 6 out. 2023b. Disponível em: https://www.nngroup.com/articles/eliza-effect-ai/?utm_source=Alertbo

Efeito Eliza: implicações da projeção antropomórfica à integridade psíquica diante da utilização de *chatbots* para cuidados com a saúde mental

x&utm_campaign=bf3e0c4d50EMAIL_CAMPAIGN_2020_11_12_08_52_COPY_01&utm_medium=email&utm_term=0_7f29a2b335-bf3e0c4d50-40700174. Acesso em: 9 jul. 2024.

STAIR, Ralph M.; REYNOLDS, George W. **Princípios de sistemas de informação**. 9. ed. São Paulo: Cengage Learning, 2011.

STOCO, Rui. **Tratado de responsabilidade civil**. 7. ed. São Paulo: Revista dos Tribunais, 2007.

SZANIAWSKI, Elimar. **Direitos de personalidade e sua tutela**. São Paulo: Revista dos Tribunais, 2002.

TENA, Lucimara Plaza. **Da ficção à realidade jurídica: a pessoa artificial e os direitos da personalidade**. São Paulo: Dialética, 2023.

TEPEDINO, Gustavo. **Temas de direito civil**. 3. ed. Rio de Janeiro: Renovar, 2004.

TOBBIN, Raíssa Arantes; CARDIN, Valéria Silva Galdino. Tecnologias vestíveis e capitalismo de vigilância: do compartilhamento de dados sobre saúde e a proteção dos direitos da personalidade. **Revista de Direito, Governança e Novas Tecnologias**, v. 7, n. 1, p. 126-147, 2021.

UNIÃO EUROPEIA (UE). Parlamento Europeu. **Disposições de Direito Civil sobre Robótica Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL))**. 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.pdf. Acesso em: 6 jul. 2024.

VARELLA, Flávia Florentino; BONALDO, Rodrigo Bragio. Negociando autoridades, construindo saberes: a historiografia digital e colaborativa no projeto Teoria da História na Wikipédia. **Revista Brasileira de História**, São Paulo, v. 40, n. 85, p. 147-170, 2020.

VUI, Chang Sim; SOON, Gan Kim; ON, Chin Kim; ALFRED, Rayner; ANTHONY, Patricia. A review of stock market prediction with artificial neural network (ANN). In: INTERNATIONAL CONFERENCE ON CONTROL SYSTEM, COMPUTING AND ENGINEERING, 2013. **Anais** [...]. IEEE, 2013. Disponível em: <https://ieeexplore.ieee.org/xpl/conhome/6709841/proceeding>. Acesso em: 6 jun. 2024.

WALKER, Lauren. Belgian man dies by suicide following exchanges with chatbot. **The Brussels Times**, Bruxelas, 28 Mar. 2023. Disponível em: <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt>. Acesso em: 22 set. 2024.

WEIZENBAUM, Joseph. ELIZA: a computer program for the study of natural language communication between man and machine. **Communications of the ACM**, v. 9, n. 1, p. 36-45, 1966.

WORLD HEALTH ORGANIZATION (WHO). **A New Agenda for Mental Health in the Americas**: Report of the Pan American Health Organization High-Level Commission on Mental Health and COVID-19. 2023. Washington: WHO, 2023. Disponível em: <https://iris.paho.org/handle/10665.2/57508>. Acesso em: 6 jun. 2024.

WORLD HEALTH ORGANIZATION (WHO). **The Comprehensive Mental Health Action Plan 2013-2030**. 2021. Disponível em: https://cdn.who.int/media/docs/default-source/campaigns-and-initiatives/world-mental-health-day/2021/mental_health_action_plan_flyer_member_states.pdf?sfvrsn=b420b6f1_7&download=true. Acesso em: 6 jul. 2024.

WORLD HEALTH ORGANIZATION (WHO). **Mental health**. 2022a. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>. Acesso em: 6 jun. 2024.

WORLD HEALTH ORGANIZATION (WHO). **World Mental Health Report: transforming mental health for all**. 2022b. Disponível em: <https://iris.who.int/bitstream/handle/10665/356119/9789240049338-eng.pdf?sequence=1>. Acesso em: 4 jun. 2024.

WOEBOT: The Mental Health Ally. **Google Play**, 2024. Disponível em: <https://play.google.com/store/apps/details?id=com.woebot&hl=en&pli=1>. Acesso em: 5 jul. 2024.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância**. Tradução George Schlesinger. São Paulo: Intrínseca, 2021.

NOTA DE COAUTORIA

O trabalho foi desenvolvido em nível de Doutorado em Direito, no âmbito do Programa de Pós-Graduação em Ciências Jurídicas da Universidade Cesumar (UNICESUMAR), cuja área de concentração é referente aos “Direitos da Personalidade”, como resultado de pesquisas voltadas à inteligência artificial, à proteção de dados e aos direitos da personalidade. Ambas as coautoras, RAÍSSA ARANTES TOBBIN (Doutoranda em Direito) e VALÉRIA SILVA GALDINO CARDIN (Docente e Orientadora) contribuíram para a definição e o aprofundamento do tema, bem como a escolha do aporte teórico e a escrita do artigo.

Como citar este documento:

TOBBIN, Raissa Arantes; CARDIN, Valeria Silva Galdino. Efeito Eliza: implicações da projeção antropomórfica à integridade psíquica diante da utilização de *chatbots* para cuidados com a saúde mental. **Revista Opinião Jurídica**, Fortaleza, v. 22, n. 41, p. 272-308, set./dez. 2024. Disponível em: link do artigo. Acesso em: xxxx.



Disponível em:

<https://www.redalyc.org/articulo.oa?id=633881856010>

Como citar este artigo

Número completo

Mais informações do artigo

Site da revista em redalyc.org

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe,
Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no
âmbito da iniciativa acesso aberto

Raissa Arantes Tobbin, Valeria Silva Galdino Cardin

**Efeito Eliza: Implicações da Projeção Antropomórfica à
Integridade Psíquica diante da Utilização de Chatbots
para Cuidados com a Saúde Mental**

**Eliza Effect: Implications of Anthropomorphic Projection
on Psychic Integrity in the Face of the use of Chatbots for
Mental Health Care**

**Efecto Eliza: Implicaciones de la Proyección
Antropomorfica para la Integridad Psíquica Respecto al
uso de Chatbots para el Cuidado de la Salud Mental**

Revista Opinião Jurídica

vol. 22, núm. 41, p. 272 - 308, 2024

Centro Universitário Christus,

ISSN: 1806-0420

ISSN-E: 2447-6641

DOI: [https://doi.org/](https://doi.org/10.12662/2447-6641oj.v22i41.p272-308.2024)

10.12662/2447-6641oj.v22i41.p272-308.2024