



# ÉTICA DA APLICAÇÃO DE INTELIGÊNCIAS ARTIFICIAIS E CHATBOTS NA SAÚDE MENTAL: UMA PERSPECTIVA PSICANALÍTICA

## THE ETHICS OF APPLYING ARTIFICIAL INTELLIGENCE AND CHATBOTS IN MENTAL HEALTH: A PSYCHOANALYTIC PERSPECTIVE

Paulo Victor dos Reis Silveira<sup>1</sup>

João Luiz Leitão Paravidini<sup>2</sup>

**Resumo:** Este artigo discute as implicações éticas do uso de Inteligências Artificiais (IAs) ou chatbots terapeutas na saúde mental através do método da revisão narrativa da literatura. Define-se o que são esses dispositivos, quais são seus benefícios e seus limites para a prática clínica. Em seguida, examinam-se quais problemas éticos que resultam dessas tecnologias pela perspectiva da psicanálise. Destacam-se os riscos de danos afetivos complexos, de falta de validade e confiabilidade das informações e de ausência de responsabilização dos agentes envolvidos. Por fim, apresentam-se algumas recomendações para o desenvolvimento de IAs mais éticas e as questões que se colocam nesse sentido.

**Palavras-chave:** Inteligência Artificial; Saúde Mental; Psicanálise; Ética.

**Abstract:** This paper discusses the ethical implications of the use of Artificial Intelligences (AIs) or therapist chatbots in mental health through the method of narrative literature review. It defines what these devices are, their benefits and limitations for clinical practice. Then, it examines the ethical problems that result from these technologies from perspective of psychoanalysis. It highlights the risks of complex emotional harm, lack of validity and reliability of information, and absence of accountability of the agents involved. Finally, it presents some recommendations for the development of more ethical AIs and the questions that arise in this sense.

**Keywords:** Artificial Intelligence; Mental Health; Psychoanalysis; Ethics.

### 1 Introdução

A emergência de Inteligências Artificiais (IAs) como o ChatGPT possibilitou uma expansão sem precedentes das tecnologias baseadas na interação humana com as máquinas. À medida que essas tecnologias ganham influência em diversas áreas do conhecimento, aumentam também as preocupações, tanto técnicas, que buscam explicar o funcionamento e o comportamento das IAs, quanto éticas. Em 2023, uma carta foi enviada mundialmente<sup>3</sup> pedindo a suspensão do treinamento de IAs mais avançadas que GPT-4, que é a versão 4 de um modelo de aprendizado de máquina que emprega métodos

<sup>1</sup> Graduado em Psicologia pelo Centro Universitário do Triângulo (UNITRI). Mestrando pela Universidade Federal de Uberlândia (UFU), Uberlândia-MG, Brasil. E-mail: [pvreis.silveira@gmail.com](mailto:pvreis.silveira@gmail.com)

<sup>2</sup> Doutor em Ciências da Saúde (Saúde Mental) pela Universidade Estadual de Campinas (Unicamp). Docente do Instituto de Psicologia - Universidade Federal de Uberlândia (UFU), Uberlândia-MG, Brasil. E-mail: [jparavidini@gmail.com](mailto:jparavidini@gmail.com) ; [paravidini@ufu.br](mailto:paravidini@ufu.br)



de aprendizado supervisionado e não supervisionado para compreender e produzir uma linguagem que imita a humana (Radford *et al.* 2018). Nesta, constam as assinaturas de pesquisadores dos maiores laboratórios tecnológicos do mundo, demonstrando o nível de alerta com tais avanços.

O progresso das IAs já alcançou o campo da saúde mental e mostra avanços significativos. Programas que usam IAs em conjunto com psiquiatras conseguem diagnosticar depressão e esquizofrenia com mais precisão do que os psiquiatras sozinhos. Além disso, essas IAs também podem prever as consequências dos tratamentos dessas doenças com maior precisão. As IAs também são usadas para auxiliar no trabalho com idosos que sofrem de demência, crianças do espectro autista e para ajudar pacientes a confrontarem suas alucinações auditivas (Grodniewicz; Hohol, 2023).

É importante e necessário refletir sobre os próximos passos do avanço das tecnologias com IA em relação à psicoterapia. Eles questionam se as IAs já estão ou serão capazes de realizar psicoterapia, e se isso representaria uma ruptura com os conceitos tradicionais sobre o assunto. O mais próximo desse modelo atualmente são as relações entre chatbots especialistas em saúde mental e usuários humanos (Grodniewicz; Hohol, 2023).

O presente trabalho tem como objetivo apresentar como os chatbots e IAs estão presentes no campo da saúde mental e quais possíveis implicações éticas relativas a presença desses dispositivos. Para isso, levantamos as seguintes questões: como a ética da psicanálise se articula com a ética do desenvolvimento e do uso de IAs? Como a ideia de Freud de que o analista não deve satisfazer a demanda do analisando se aplica às configurações e articulações no nível social entre humanos e IAs? Como o desenvolvimento de IAs cada vez mais capazes de simular uma consciência afeta a estrutura narcísica do sujeito?

A importância científica e social da discussão que o artigo se propõe a fazer sobre "as implicações éticas do uso de Inteligências Artificiais (IAs) ou chatbots terapeutas na saúde mental", tomando em perspectiva o que nos adverte a psicanálise quanto à ética do desejo. Desse modo, trata-se de revisão narrativa da literatura, que tem o intuito de descrever e discutir o desenvolvimento ou o "estado da arte" deste tema, sob ponto de vista teórico ou contextual, produzindo uma avaliação abrangente da literatura (Rother, 2007).



## 1.1 Chatbots terapêuticos

A inteligência artificial (IA) é uma forma inovadora de ampliar interações terapêuticas sem mediação humana. Agentes terapêuticos relacionais, baseados em alguns princípios da psicologia, como na relação terapeuta-paciente, podem estabelecer empatia, confiança e simular uma aliança terapêutica com os usuários. Evidências recentes (Lucas *et al.* 2014) mostram que as pessoas tendem a divulgar mais informações sensíveis durante as entrevistas clínicas iniciais realizadas por computadores do que quando realizadas por humanos. O teste realizado com 97 participantes compara as respostas dos usuários na interação com humanos e na interação com agentes virtuais nos quais os usuários reportaram serem mais capazes de revelarem informações sobre o que estavam sentindo e sobre tristeza com os agentes virtuais.

Uma forte aliança terapêutica pode ser formada mesmo sem contato presencial, inclusive com interações não humanas com aplicativos ou com um agente de conversação (chatbots). Apesar de a maioria dos aplicativos de saúde mental ser abandonada em poucos dias após a instalação, evidências meta-analíticas sugerem que os chatbots podem aumentar o engajamento e o prazer no cuidado de saúde mental digital (Prochaska *et al.* 2023).

Os chatbots, em sua maioria, servem como uma porta de entrada para a terapia convencional, indicando que os usuários busquem um terapeuta humano (Coghlan *et al.* 2023). Outros, entretanto, também oferecem assistência aos usuários quanto à saúde mental, passando exercícios baseados em terapia cognitivo-comportamental e emulando emoções e conexões psicológicas. Essas interações ocorrem por chat de conversação direta e são atraentes por serem facilmente acessíveis, como no download em um smartphone, e por estarem disponíveis integralmente aos usuários.

Novas formas de terapias digitais, como chatbots terapêuticos, estão mostrando resultados promissores no tratamento de diversos tipos de problemas de saúde mental. Um estudo de Suharwardy *et al.* (2023) mostrou que o chatbot Woebot foi eficaz na redução dos sintomas de depressão, mesmo sem supervisão humana. Em outro estudo, dependentes químicos que interagiram com agentes terapeutas apresentaram redução do uso de substâncias como álcool e drogas. Além disso, chatbots terapêuticos são mais efetivos do que a psicoeducação digital.

O Woebot é um chatbot terapêutico que simula diversos elementos de uma sessão de terapia real. Ele usa técnicas de rapport, como fazer perguntas abertas e demonstrar



interesse, para estabelecer uma conexão com o usuário. Também confirma se entendeu direito o que o usuário disse, normaliza a experiência do usuário e demonstra linguagem empática. Além disso, o Woebot sugere um tipo de aliança terapêutica e trabalha psicoeducação com o usuário. Por fim, o Woebot aplica conceitos chave de terapia cognitiva, convidando o usuário a refletir e reformular certos padrões compreendidos durante a conversa (Grodniewiecz; Hohol, 2023).

Apesar dos numerosos benefícios da tecnologia de inteligência artificial (IA) nos cuidados de saúde mental, também existem algumas limitações que precisam ser abordadas. Uma das principais limitações é a falta de dados sobre certos transtornos mentais. As ferramentas de IA dependem de grandes quantidades de dados para fornecer diagnósticos precisos e confiáveis, e a falta de dados pode limitar a eficácia dessas ferramentas. Além disso, a falta de transparência e responsabilização sobre o uso de IA nos cuidados de saúde mental também é uma preocupação. Os pacientes precisam estar cientes de como seus dados estão sendo usados e ter a capacidade de controlar o uso deles (Rana; Singh, 2023).

Outra limitação dos cuidados de saúde mental baseados em IA é o potencial de viés algorítmico. As ferramentas de IA são tão eficazes quanto os dados em que são treinadas. Portanto, se os dados forem tendenciosos, os resultados também serão tendenciosos, levando a diagnósticos e recomendações de tratamento incorretos. As considerações éticas em torno do uso de IA nos cuidados de saúde mental precisam ser abordadas. Isso inclui questões como privacidade, consentimento informado e o potencial da IA para substituir a interação humana (Rana; Singh, 2023).

## 1.2 Ética e Chatbots

Segundo Coghlan *et al.* (2023), para que um chatbot seja funcional e aceito pelo usuário, é necessário que ele seja seguro, satisfaça as demandas de maneira efetiva e que haja uma forma de feedback com uma equipe de desenvolvedores envolvidos. Entre os aplicativos que falharam nessas exigências fundamentais dos usuários estão os que, de forma prejudicial e por vezes antiéticos, se tornaram ameaças por serem ofensivos, causarem danos emocionais, confundirem as perguntas dos usuários e não cumprirem com o que prometiam entregar de vantagens.

Outros riscos podem ser mais graves e até mesmo devem ser discutidos com maior rigor. Na Bélgica, um homem cometeu suicídio após seis semanas de conversação intensa



com o chatbot ELIZA (Walker, 2023). O programa simulava uma interação que uma pessoa teria com um terapeuta da abordagem centrada na pessoa de Carl Rogers, sendo capaz de enganar usuários a pensarem que realmente estavam falando com um terapeuta humano. Aqui se levanta a primeira questão ética, pois simular algo é uma forma de enganar e quanto mais enganado é o usuário mais efetivo é o programa (Grodiewicz; Hohol, 2023).

Tekin (2023) levanta três pontos importantes de preocupação ética no uso de chatbots na saúde mental. O primeiro é o que o autor denomina “o bot não é um terapeuta”. Para ele, é incorreto chamar o que os chatbots fazem de “terapia”, nem podemos chamar os chatbots de “terapeutas”. Os chatbots não podem prover o mesmo nível de cuidado e atenção que um terapeuta humano, além de poder minar a importância da construção de confiança entre os pacientes e os profissionais de saúde que poderiam mal interpretar o papel do psicólogo de modo geral.

O segundo ponto levantado por Tekin (2023) é a “presunção de rastreabilidade”, que assume o ponto de vista de que os usuários serão honestos e verdadeiros com o seu chatbot. No entanto, pessoas que tem conhecimento sobre dados privados e como esses dados podem “vazar”, relutam em revelar informações sensíveis aos chatbots, e podem não se sentir confortáveis. Em consequência disso, o chatbot criaria diagnósticos imprecisos ou recomendações de tratamento ineficazes. Existe também uma preocupação com a privacidade dos dados, pois os chatbots coletam informações pessoais sensíveis dos usuários, e há um risco de que essas informações sejam mal utilizadas ou compartilhadas sem o consentimento do usuário, como visto em empresas como Facebook com a manipulação ou vendas de informações que se tornou preocupação pública amplamente divulgada.

Ainda no que tange a esse segundo ponto, a pesquisa de Lucas *et al.* (2014) e Tekin (2023) apresentam pontos de vista distintos, porém complementares. Lucas *et al.* (2023) sugerem que os usuários tendem a revelar mais informações sensíveis para agentes virtuais do que para humanos. Por outro lado, Tekin (2023) argumenta que a presunção de que os usuários serão honestos com os chatbots é um problema ético, especialmente se os usuários estiverem cientes dos riscos de privacidade associados à divulgação de informações sensíveis. O levantamento ético de Tekin (2023) se dá pela crescente onda de vazamentos de dados, o que não foi observado na interação das entrevistas iniciais do experimento realizado por Lucas *et al.* (2014) e pode apontar para uma ampliação na confiabilidade da interação caso haja clareza, transparência e segurança sobre como os



dados da interação poderão ser utilizados ou conforme a segurança de dados vai se aperfeiçoando.

O terceiro ponto de preocupação apresentado por Tekin (2023) é com a “lacuna de evidência”, que se refere à falta de evidência empírica que sustente a efetividade dos chatbots no diagnóstico e tratamento de condições de saúde mental. Embora alguns estudos tenham mostrado resultados promissores, ainda há uma necessidade de mais pesquisas rigorosas para determinar se os chatbots podem prover diagnósticos e tratamentos eficazes e seguros para condições de saúde mental. Sem essa evidência, é difícil justificar o uso de chatbots no cuidado da saúde mental, e há um risco de que os pacientes possam ser prejudicados ao dependerem de tratamentos não comprovados ou ineficazes.

Segundo Sedlakova e Trachsel (2022), há o problema de não haver um consenso de precauções éticas nem guias e critérios no desenvolvimento de chatbots. Também não há uma normativa ou órgão para analisar se tais programas estão sendo desenvolvidos de forma ética. Além disso, há uma preocupação, pois um chatbot não tem o mesmo nível de responsabilidade e deveres de um terapeuta humano, uma vez que uma IA não é um “sujeito moral” e não há regulamentações legais claras quanto aos maus resultados dessa relação.

Existe preocupação com a prevenção de danos, um campo da ética em tecnologia que lida com o mal funcionamento, operação de forma não prevista ou roubo de informações por hackers ou monitoramento não autorizado (Fiske; Henningsen; Buyx, 2019). Além disso, deve-se ter cuidado com órgãos governamentais ou serviços de saúde substituírem os serviços de saúde por ferramentas mais baratas como chatbots terapeutas, resultando em menor disponibilidade os recursos já existentes em detrimento de tecnologias de IA.

Fiske, Henningsen e Buyx (2019) apontam que toda tecnologia de IA trabalha com algoritmos e bases de dados e que ambos estão sujeitos a problemas éticos, como viés humanos que reproduzem e reforçam formas de desigualdade social já existentes, podendo gerar chatbots racistas e sexistas, e a falta de transparência exigida por conta da competitividade do mercado dificulta formas de atuar pontualmente no problema e identificar possíveis problemas, deixando toda responsabilidade nas mãos dos próprios programadores.



### 1.3 Ética na Psicanálise

A teoria psicanalítica de Freud e Lacan pode acrescentar alguma densidade na discussão quanto a ética tanto na produção de IAs quanto no desenvolvimento de IAs mais éticas pois pensa o laço social e o campo social a partir da relação que, por extensão se aplica a relação sujeito e máquinas. Além disso é possível questionar as normas impostas pela cultura, sociedade e do discurso dominante dando espaço para pensar através de uma ética da singularidade.

Desde sua origem, a psicanálise tem abordado a ética como forma de leitura dos fenômenos socioculturais. Essa leitura, segundo Rosa, Carignato e Berta (2006), evidencia o modo de laço social que constitui a cultura, seja sob forma de consumo, lucro ou sofrimento. A psicanálise, segundo Palumbo (2016), pode contribuir para uma compreensão social mais complexa, pois o sujeito não é apenas um produto da sociedade, mas também um sujeito desejante que pode se posicionar de forma crítica em relação às normas e valores sociais. A ética da psicanálise reintroduz, portanto, o laço social, “às avessas”, pois convida o sujeito a assumir seu desejo e refletir sobre sua participação subjetiva no laço social.

Freud (1919/1996, p.214) ao abordar a ética na psicanálise, defende que a terapia analítica não deve satisfazer os desejos do paciente, mas sim mantê-lo em um estado de privação ou abstinência, dado que a frustração constituía um estímulo para o paciente mudar, e a satisfação das necessidades podia funcionar como substituto dos sintomas e impedir sua elaboração e transformação. Além disso, a terapia analítica não deve impor ao paciente os ideais do analista, nem tentar moldá-lo à sua imagem. Isso seria uma forma de violência e de autoritarismo, que impediria o paciente de desenvolver sua própria personalidade e autonomia. Esse modelo serve também para o social, pois, ainda segundo Freud (1921/1996, p.81), desde o início a psicologia individual é ao mesmo tempo uma psicologia social.

Freud (1913/1990a, p.164) afirma que "a extraordinária diversidade das constelações psíquicas envolvidas, a plasticidade de todos os processos mentais e a riqueza dos fatores determinantes opõem-se a qualquer mecanização da técnica" renunciando às regulamentações e regras em prol das manifestações singulares de cada indivíduo, sendo impossível qualquer tentativa de formulação de uma técnica única e verdadeira. O objetivo da psicanálise deve ser o de auxiliar o sujeito a se reconciliar com seu desejo, e encontrar um modo de viver com ele de forma satisfatória.



Lacan (1959-60/1992, p. 314) retoma essa perspectiva freudiana ao afirmar que “a ética da psicanálise é a ética que convém à nossa ciência, na medida em que ela é a ciência do sujeito”. Para Lacan, o sujeito é aquele que se constitui no campo da linguagem, marcado pela falta e pelo desejo. O desejo do sujeito não é um simples apetite ou uma necessidade biológica, mas uma busca incessante de um objeto perdido que nunca se encontra. O desejo é o que move o sujeito em sua análise, levando-o a confrontar-se com o inconsciente, com o Outro e com o gozo. O gozo é uma forma de satisfação paradoxal e dolorosa, que ultrapassa o princípio do prazer e que implica uma transgressão dos limites impostos pela lei simbólica.

O gozo é o que resiste à análise, pois é o que escapa à significação e à interpretação. A ética da psicanálise, segundo Lacan (1959-60/1992, p. 366), consiste em “não ceder em seu desejo”, ou seja, em não renunciar ao seu desejo em nome de um ideal ou de uma norma moral. A ética da psicanálise visa a uma sublimação do gozo em um ato criativo e singular, que afirme a diferença do sujeito. Há, portanto, uma ética do desejo que se diferencia da ética social.

Segundo Possati (2023), a ética não pode ser reduzida a um código e, ao mesmo tempo, conforme a tecnologia avança, mais se faz necessário aplicar capacidades de análises éticas aos sistemas de IA. A discussão da ética deve partir do ponto de vista do design e responder aos três níveis: ética no design, ética por design e ética para design. A ética por design é uma forma de integrar os princípios e valores éticos na concepção, implementação e avaliação de sistemas de IA. A ideia é que a ética não seja vista como um obstáculo ou uma restrição, mas como um elemento essencial e orientador do processo de desenvolvimento de IA.

Uma perspectiva de pesquisa inspirada na sociologia e na psicanálise aprimora a análise dos problemas clássicos da IA, pois leva em consideração os aspectos humanos e sociais envolvidos. A psicanálise permite levantar um novo ponto de vista sobre as questões éticas levantadas pela IA de uma forma mais relacional, levando em consideração os interesses e perspectivas não apenas dos agentes, mas também dos usuários (Possati, 2023). Por essa ótica é possível analisar os problemas clássicos da IA não apenas sob a ótica de um único sistema, mas como parte de uma comunidade composta por humanos e não-humanos. Os problemas éticos são vistos em seu contexto histórico e social. Além disso, a psicanálise nos ajuda a aceitar e compreender os limites da responsabilidade, reconhecendo que existem situações em que não é possível eliminar completamente as responsabilidades e problemas. Se um sistema de IA demonstra um



comportamento discriminatório em relação às mulheres, esse viés deve ser analisado como um fenômeno coletivo, um “habitus” que a máquina assimilou de um contexto humano-não-humano. Para explicar esse viés, seria necessário reconstruir a comunicação inconsciente do habitus para a máquina em um contexto social específico. E para corrigir esse viés, seria necessário criar novas condições de comunicação nesse campo social.

## 2 Recomendações éticas no desenvolvimento de chatbots

Coghlan *et al.* (2023) afirma que para que um chatbot seja funcional e aceito pelo usuário, é necessário que ele siga os princípios éticos de não maleficência, beneficência, autonomia, justiça e explicabilidade (transparência e responsabilidade). Para isso, eles devem avaliar os riscos e benefícios do uso de chatbots, buscar e divulgar evidências que sustentem sua eficácia e segurança, respeitar a privacidade e o consentimento dos usuários, garantir a qualidade e a acessibilidade dos serviços e monitorar e avaliar os resultados dos chatbots. Também devem considerar a possibilidade da revelação de crimes às autoridades, caso detectem emergências ou risco iminente para os usuários ou terceiros. Além disso, devem buscar uma maior participação e consulta dos usuários na criação e regulação dos chatbots, respeitando a diversidade, as particularidades e as experiências dos afetados.

Fiske, Henningsen e Buyx (2019) recomendam a criação de orientações claras sobre a regulação, a supervisão e o consentimento da IA incorporada, bem como que se treine os profissionais de saúde e se consulte os usuários sobre o uso da IA. A utilização da IA na saúde mental deve ser uma adição e não substituição dos recursos presentes nas instâncias da saúde mental. Ainda é necessário que se faça pesquisa sobre os impactos diretos e indiretos da IA incorporada na relação terapêutica, nas outras relações humanas e na autoconsciência, agenciamento e identidade dos sujeitos.

Tawfeeq, Awqati e Jasim (2023) afirmam que para proteger o usuário é necessário a regulamentação do uso de sistemas de IA como o ChatGPT, estabelecendo regras para priorizar o bem-estar do usuário, protegendo-os de possíveis danos causados pela conexão emocional ou dependência do sistema. Também devem garantir transparência e responsabilidade dos desenvolvedores em relação ao design e treinamento dos sistemas de IA, responsabilizando-os por qualquer conteúdo prejudicial ou inadequado produzido pelo chatbot. Outro ponto fundamental é o aumento da educação e da conscientização entre desenvolvedores, usuários e o público em geral sobre as implicações éticas dos



chatbots em geral. Outro ponto importante é a necessidade de estabelecer governança e regulamentação adequadas para os sistemas de IA, incluindo supervisão regulatória e adesão a padrões éticos. Por fim, é importante integrar questões éticas relevantes aos chatbots com as estruturas éticas já existentes e diretrizes para pesquisa e uso de IA.

Mas estabelecer normativas a serem seguidas é o suficiente para manter uma prática ética? Um estudo controlado sobre as decisões individuais dos engenheiros de software, revela que as diretrizes têm um impacto quase nulo no comportamento dos profissionais da área de tecnologia (Hagendorff, 2020). Os participantes foram submetidos a onze cenários de decisão ética relacionados ao software que abrangiam temas como responsabilidade de reportar problemas e erros, coleta de dados do usuário, propriedade intelectual, qualidade do código, honestidade com o cliente e gestão de tempo e pessoal. O estudo envolveu 63 estudantes e 105 desenvolvedores de software profissionais. Os pesquisadores avaliaram se o código de ética dos profissionais da computação exercia algum impacto nas decisões éticas em seis desses cenários e concluiu-se que não houve diferença estatisticamente significativa nas respostas entre os indivíduos que viram ou não o código de ética, tanto para estudantes quanto para profissionais.

A ética enfrenta limitações tanto no nível individual quanto no social como um todo (Hagendorff, 2020). No mercado competitivo, muitas empresas buscam lucrar com a IA em diversas aplicações, seguindo uma lógica econômica que não é condicionada por uma ética baseada em valores ou princípios. Os engenheiros e desenvolvedores, por sua vez, não recebem formação nem apoio para lidar com as questões éticas. No contexto empresarial, a velocidade é essencial e ignorar a ética é o caminho mais fácil. Assim, a prática de desenvolvimento, implementação e uso de aplicações de IA tem pouco a ver com os valores e princípios propostos pela ética. Muito dinheiro é investido no desenvolvimento e uso comercial de sistemas baseados em aprendizado de máquina, enquanto as considerações éticas são usadas principalmente para fins de relações públicas.

Para uma ética da IA eficaz, é preciso ir além das diretrizes de verificação e adotar uma abordagem sensível às situações, que valoriza as virtudes e as disposições pessoais, as ampliações do conhecimento, a autonomia responsável e a liberdade de ação dos atores morais, não buscando generalizar os casos sob princípios individuais, mas se adaptando às situações específicas e às configurações técnicas. Essa abordagem requer um equilíbrio entre o foco nos detalhes tecnológicos dos métodos e tecnologias de IA e do aprendizado



de máquina, para aproximar a ética dos discursos técnicos, e o foco nos aspectos sociais e pessoais, para emancipar os atores morais de rotinas problemáticas e promover a responsabilidade individual. A ética da IA então lida menos com a IA em si, do que com formas de desvio ou distanciamento dos programadores com o que é produzido (Hagendorff, 2020).

Nesse ponto, além de entender e analisar as diretrizes éticas que orientam o desenvolvimento de IAs, é importante analisar quais aspectos impactam a cultura de modo geral, para evidenciar tais aspectos com maior profundidade e transparecer o modo que tais laços operam no sujeito social. Esse laço é determinante para entender a relação humano e não humano e pensar na ética só pelo ponto de vista da tecnologia implica em pensar um programa neutro que não se implica na relação e não é afetado pelo sujeito, o que não é possível pensar pela ótica da psicanálise.

Pensar na educação ética dos engenheiros de software é pensar na educação do ponto de vista da ética do desejo. Para Costardi e Endo (2013), a educação se compromete com um projeto civilizatório e faz uma demanda moral ao sujeito, que deve se inserir no âmbito dos ideais coletivos sem abdicar de sua singularidade. Essa inserção implica um custo para o desejo do sujeito, que deve renunciar a parte de sua satisfação. A ética do desejo, advinda da psicanálise, não se opõe a esse processo, mas o problematiza, na medida em que reconhece as resistências das singularidades à proposta homogeneizadora de uma educação. Tais resistências são expressões do sujeito do inconsciente, que escapa às técnicas didáticas e aos saberes especializados que se dirigem à educação, seja ela realizada dentro da escola ou no sentido mais amplo da experiência social. A ética da psicanálise afeta a educação, na medida em que ela tem a ver com a capacidade de lidar com a maldade fundamental por outras vias que não a bondade, de suportar um furo no saber que pretendia aliar meios e fins, para colocar em jogo um saber inédito e particular que, às vezes, precisa ser inventado pelos sujeitos que estão implicados na empreitada educativa.

Na perspectiva ética de Lacan (1959/60) o desejo do sujeito é o desejo do Outro e, ao entrar na relação com a lei, a tradição e a moral, tem que renunciar a parte de seu gozo, ou seja, de sua satisfação pulsional, para se identificar com os ideais coletivos. Essa renúncia é a castração simbólica, sendo ela a condição para o acesso ao campo do desejo. É importante reconhecer que há um resto de gozo que escapa à castração simbólica e que se manifesta como um excesso, uma falta ou uma transgressão. Esse resto ou falta é o que Lacan chama de objeto *a* que é o objeto causa do desejo, mas que nunca pode ser



plenamente alcançado ou satisfeito. O objeto *a* é o ponto de articulação entre o simbólico e o real, entre a lei e o gozo. A ética da psicanálise consiste em não ceder em seu desejo, ou seja, em não renunciar ao objeto *a* em nome de uma adaptação às normas sociais ou de uma ilusão de completude. A ética da psicanálise implica em assumir a divisão do sujeito, a falta no Outro e a impossibilidade de um saber absoluto sobre o desejo.

Considerando essa perspectiva lacaniana do desejo, a produção de IAs mais éticas, deve-se colocar em evidência as incongruências das pretensões de universalidade, racionalidade e transparência que muitas vezes são atribuídas a produção ética de tais tecnologias. As IAs não são apenas sistemas lógicos e computacionais, mas também podem ser compreendidas como sistemas simbólicos e afetivos, que se relacionam com os sujeitos humanos e com os seus desejos. Os chatbots não podem ignorar as dimensões do inconsciente, do gozo e da fantasia que estão presentes nas interações humanas. É necessário pensar em como essas tecnologias serão capazes de lidar com a complexidade e a ambiguidade das situações éticas, sem se reduzirem a algoritmos ou protocolos pré-definidos e em como ou se serão responsáveis pelos seus atos e pelos seus efeitos, sem se eximirem das consequências de suas decisões.

Possati (2023) propõe que o desenvolvimento de IA deve ser baseado na responsabilidade de agentes (usuários, desenvolvedores, programadores, designers, proprietários e software) e pacientes (que usam a IA e interagem com ela). Para ele, não pode haver responsabilidade real sem a relação entre o agente e o paciente, pois o agente não pode agir eticamente sem existir a demanda do paciente, independentemente de todos os princípios morais ou virtudes possíveis. Abordar a dimensão ética requer considerar os indivíduos que concebem, edificam e aplicam a inteligência artificial em sua comunidade. A ética, entretanto, não se alinha com a criação de regulamentos globalmente aplicáveis nem com obrigações puramente subjetivas. Em vez disso, no contexto das interações entre seres humanos e IA, implica compreender os sistemas de IA como sujeitos e objetos de avaliação ética em meio a uma rede complexa de relações entre indivíduos e não humanos.

Oliveira e Corrêa (2023) afirmam que a virtualidade tecnológica afeta e modifica a formação dos laços sociais. A nova realidade descarta aspectos perturbadores e indesejados e o lembrete virtual de uma relação sem corpo, presente na relação por uma IA realça e evidencia ao mesmo tempo os laços possíveis e impossíveis dessa relação humano e não humano, ou como proposto anteriormente, entre agentes e pacientes.



Pensar nos problemas gerados pela expansão das tecnologias de IAs que compreendem o campo da saúde mental pelo viés psicanalítico possibilita inflexões acerca da natureza da relação e como os discursos podem se articular com as utilizações dessas tecnologias. A psicanálise contribui nas discussões das naturezas das relações e, portanto, é fundamental na discussão do presente trabalho.

### 3 Conclusão

Neste trabalho, buscamos analisar as implicações éticas do uso de chatbots e inteligências artificiais (IAs) no campo da saúde mental, a partir da perspectiva da psicanálise. Consideramos que a psicanálise pode oferecer uma contribuição relevante para compreender as relações entre humanos e IAs, bem como os impactos subjetivos e morais originados dessa relação.

Retomamos a ideia de Freud de que o analista não deve satisfazer a vontade do analisando e a ampliamos para as configurações e articulações no nível social entre humanos e IAs. Argumentamos que essa lógica se faz presente na relação humano e não humano, pois a IA tanto pode atender às demandas do usuário, oferecendo soluções aos seus conteúdos psíquicos, angústias e dores, mas também pode frustrar ou desafiar o usuário, levando-o a questionar ou modificar seus padrões de pensamento e comportamento. Exemplificamos essa lógica com casos concretos de chatbots que utilizam técnicas psicoterapêuticas, como o Woebot e a ELIZA, e discutimos como eles podem afetar não só a relação do sujeito com a IA, mas também com os outros laços sociais do sujeito. Questionamos quais os prejuízos do acesso a uma ferramenta que está em total disponibilidade ao usuário atendendo a todas as suas demandas. Seria a IA capaz de encarnar a demanda ao Outro, satisfazendo assim para sustentar a obediência aos ideais sociais?

Uma das questões que emerge é como se teoriza a responsabilidade do criador (desenvolvedor, proprietário, designer e software) das possíveis falhas e 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 manter um órgão responsável por



estabelecer tais critérios e com base em qual ponto de vista são algumas das possíveis contradições evidenciadas durante a escrita deste trabalho.

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.

Não é possível pensar numa ética que não envolva o sujeito na própria relação ética a qual ele se articula e não é possível pensar numa neutralidade artificial total quando se depara com programadores humanos e relação com um sujeito não neutro que é o humano. Não há isenção numa relação, seja ela artificial ou não.

## Referências

- COGHLAN, S.; LEINS, K.; SHELDERICK, S.; CHEONG, M.; GOODING, P.; D'AIFONSO, S. To chat or bot to chat: Ethical issues with using chatbots in mental health. **Digital Health**, Londres, v. 9, n.1 p. 1-10, jun. 2023. Disponível em: <https://doi.org/10.1177/20552076231183542>. Acesso em: 22 set. 2023.
- CORDEIRO, A. M.; OLIVEIRA, G. M. de; RENTERÍA, J. M.; GUIMARÃES, C. A. Revisão sistemática: uma revisão narrativa. **Revista do Colégio Brasileiro de Cirurgiões**, Rio de Janeiro, v. 34, n. 6, p. 428-431, nov. 2007. Disponível em: <https://doi.org/10.1590/S0100-69912007000600012>. Acesso em: 29 set. 2023.
- COSTARDI, G. G.; ENDO, P. C. Ética da psicanálise, educação e civilização. **Estilos clin.**, São Paulo, v. 18, n. 2, p. 327-341, ago. 2013. Disponível em [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1415-71282013000200008&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1415-71282013000200008&lng=pt&nrm=iso). Acesso em: 29 set. 2023.
- ELIAS, C. S. R.; SILVA, L. A.; MARTINS, M. T.; RAMOS, N. A.; SOUZA, M. G.; HIPÓLITO, R. L. Quando chega o fim?: uma revisão narrativa sobre terminalidade do período escolar para alunos deficientes mentais. **Rev. Eletrônica Saúde Mental e drogas**, Ribeirão Preto, v. 8, n. 1, p. 48-53, abr. 2012. Disponível em: [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1806-69762012000100008&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1806-69762012000100008&lng=pt&nrm=iso). Acesso em: 30 jan. 2024
- FISKE, A.; HENNINGSEN, P.; BUYX, A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy.



**Journal of medical Internet research**, Toronto v. 21, n. 5, p. e13216, maio. 2019. Disponível em: <https://doi.org/10.2196/13216>. Acesso em: 22 set. 2023.

FREUD, S. Sobre o início do tratamento: novas recomendações sobre a técnica da Psicanálise. In: FREUD, S. O caso Schreber, artigos sobre técnicas e outros trabalhos. **Edição standard brasileira das obras psicológicas completas de Sigmund Freud Vol. 12**. Tradução de J. Salomão. Rio de Janeiro: Imago, 1990a. p. 74-89. (Trabalho original publicado em 1913).

FREUD, S. Linhas de progresso na terapia psicanalítica. In: FREUD, S. História de uma neurose infantil e outros trabalhos. **Edição Standard Brasileira das Obras Completas de Sigmund Freud Vol. 17**. Tradução de J. Salomão. Rio de Janeiro: Imago, 1996. p. 98-104. (Trabalho original publicado em 1919).

FREUD, S. Psicologia de grupo e análise do ego. In: FREUD, S. Além do princípio do prazer. **Edição Standard Brasileira das Obras Completas de Sigmund Freud Vol. 18**. Tradução de J. Salomão. Rio de Janeiro: Imago, 1996. p. 44-90. (Trabalho original publicado em 1921).

GRODNIEWICZ, J. P.; HOHOL, M. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. **Frontiers in Psychiatry**, Laussane, v. 14, n.1 p. 01, maio. 2023. Disponível em: <https://doi.org/10.3389/fpsy.2023.1190084>. Acesso em: 22 set. 2023.

HAGENDORFF, T. The ethics of AI ethics: an evaluation of guidelines. **Minds & Machines**, Cham v. 30, n.1 p. 99-120, fev. 2020. Disponível em: <https://doi.org/10.1007/s11023-020-09517-8>. Acesso em: 29 set. 2023.

LACAN, J. **Seminário 7 – A Ética da psicanálise**. Tradução de A. V. Milani. Rio de Janeiro: Jorge Zahar Editor, 1997. (Trabalho original publicado em 1959-60).

LUCAS, G. M.; GRATCH, J.; KING, A.; MORENCY, L-P. It's only a computer: virtual humans increase willingness to disclose. **Computers in Human Behavior**, Amsterdam, v. 37, n.1, p. 94-100, ago. 2014. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0747563214002647?via%3Dihub>. Acesso em: 22 set. 2023.

OLIVEIRA, G. D. F.; CORREA, H. C. S. Entre encontros faltosos e excessivos: laços amorosos e uso de tecnologias para pensar o sujeito. **Tempo psicanalítico**, Rio de Janeiro, v. 55, n. 1, p. 32-56, jun. 2023. Disponível em [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S0101-48382023000100002&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0101-48382023000100002&lng=pt&nrm=iso). Acesso em: 27 set. 2023.

POSSATI, L. M. **Unconscious Networks: Philosophy, Psychoanalysis, and Artificial Intelligence**. New York, NY: Routledge, 2023. Disponível em: <https://doi.org/10.4324/9781003345572>. Acesso em: 23 set. 2023.

PROCHASKA, J. J.; VOGEL, E. A.; CHIENG, A.; BAIOCCHI, M.; PAJARITO, S.; PIRNER, M. DARCY, A.; ROBINSON, A. 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, n1. P. 01, abr. 2023. Disponível em: <https://doi.org/10.1016/j.cct.2023.107125>. Acesso em: 22 set. 2023.

RADFORD, A.; NARASIMHAN, K.; SALIMANS, T.; SUTSKEVER, I. Improving language understanding by generative pre-training. **Preprint**. 1-12. 2018. Disponível em: <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>. Acesso em: 25 jan. 2024.



RANA, U.; SINGH, R. The Role of Artificial Intelligence in Mental Health Care. **SocArXiv Papers**, College Park. p.1. set. 2023 Disponível em: <https://doi.org/10.31235/osf.io/r4umy> . Acesso em: 22 set. 2023.

ROSA, M. D.; CARIGNATO, T. T.; BERTA, S. L. Ética e política: a psicanálise diante da realidade, dos ideais e das violências contemporâneos. **Ágora: Estudos em Teoria Psicanalítica**, Vitória, v. 9, n. 1, p. 35-48, jan. 2006. Disponível em: <https://doi.org/10.1590/S1516-14982006000100003>. Acesso em: 22 set. 2023.

ROTHER, E. T. Revisão sistemática X revisão narrativa. **Acta Paulista de Enfermagem**, São Paulo. v. 20, n. 2, p. v-vi, abr. 2007. Disponível em: <https://www.scielo.br/j/ape/a/z7zZ4Z4GwYV6FR7S9FHTByr/#>. Acesso em: 09 fev. 2024.

SEDLAKOVA, J.; TRACHSEL, M. Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? **The American Journal of Bioethics**, Londres, v. 23, n .5, p .4 -13, maio. 2023. Disponível em: <https://doi.org/10.1080/15265161.2022.2048739>.

SUHARWARDY, S.; RAMACHANDRAN. M.; LEONARD S. A.; GUNASEELAN A.; LYELL D. J.; DARCY A.; ROBINSON A.; JUDY A. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. **AJOG Global Reports**, Amsterdam, v. 3, n. 3, p. 1, ago. 2023. Disponível em: <https://doi.org/10.1016/j.xagr.2023.100165>. Acesso em: 29 set. 2023.

TAWFEEQ, T. M.; AWQATI, A. J.; JASIM, Y. A. The Ethical Implications of ChatGPT AI Chatbot: A Review. **Journal of Modern Computing and Engineering Research**, Mosul, v.2023, n.1 p .49 -57, jul. 2023. Disponível em: <https://jmcer.org/research/the-ethical-implications-of-chatgpt-ai-chatbot-a-review/>. Acesso em: 25 set. 2023.

TEKIN, Ş. Ethical issues surrounding artificial intelligence technologies in mental health: psychotherapy chatbots In: ROBSON, G. J.; TSOU, J. (ed.). **Technology ethics: A philosophical introduction and readings**. Londres: Routledge, 2023. p. 152 -159. Disponível em: <https://doi.org/10.4324/9781003189466>.

WALKER, L. 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. 2023.

**Recebido em:** 12 de novembro de 2023.

**Aceito em:** 24 de janeiro de 2024.