

A inteligência artificial pode substituir o psicanalista?

A crítica de Terrence Deacon à teoria computacional sobre a mente

Eduardo Zaidhaft,¹ Rio de Janeiro
Ramon Reis,² Rio de Janeiro
Monah Winograd,³ Rio de Janeiro

Resumo: A inteligência artificial (IA) generativa, com sua capacidade de executar tarefas léxicas antes restritas ao comportamento humano, reatualiza o debate sobre a capacidade de computadores replicarem o pensamento, propondo questões para a psicanálise, tanto teóricas quanto clínicas. Os autores examinam possibilidades e limitações de uma prática analítica mediada por algoritmos, considerando a questão: a IA pode substituir o psicanalista? Argumentam, criticamente à teoria computacional para o problema da mente, que a psicanálise seria inviável por meio de mecanismos de IA, desde que sua prática esteja fundamentada nos aspectos semântico, intencional e normativo das representações mentais, em contraste com abordagens que priorizam analisar o caráter sintático da associação livre e dos processos inconscientes.

Palavras-chave: Sigmund Freud, tecnologias digitais, computacionalismo, emergentismo, Terrence Deacon

- 1 Membro provisório da Sociedade Brasileira de Psicanálise do Rio de Janeiro (SBPRJ). Professor da Faculdade de Medicina da Universidade Estácio de Sá (Unesa). Programa de Pós-Graduação em Psicologia Clínica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).
- 2 Programa de Pós-Graduação em Psicologia Clínica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).
- 3 Professora associada do Programa de Pós-Graduação em Psicologia Clínica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Coordenadora do Laboratório de Pesquisas Avançadas em Psicanálise e Subjetividade (Lapsu) e do Laboratório de Humanidades Digitais da PUC-Rio.

Introdução

Uma vertente investigativa no diálogo entre psicanálise e tecnologias digitais aplica conceitos freudianos ao estudo da inteligência artificial (IA) e às interações com dispositivos tecnológicos. Turkle (1988) sugere uma aliança entre psicanálise e IA, ancorada na ideia de uma agência fragmentada e descentralizada, análoga ao sujeito freudiano. Souza Leite (2022), por sua vez, explora uma perspectiva distinta ao comparar parcialmente a capacidade das inteligências artificiais com o conceito de construção em Freud e com a função-alfa de Bion. Segundo o autor, os algoritmos seriam capazes de inferir comportamentos conscientes a partir de dados inconscientes ou não psíquicos, traçando uma aproximação entre processos computacionais e dinâmicas psíquicas.

Freud utilizou diversas concepções mecanicistas para descrever meta-psicologicamente o psiquismo, incluindo o modelo topográfico, os pontos de vista econômico e dinâmico, bem como metáforas elétrico-hidráulicas para abordar a pulsão. Contudo, alguns autores, como Ruz (2019), sugerem que ele se afastou dessa perspectiva mecanicista em direção a uma abordagem que poderia ser descrita como “vitalista negativa”, especialmente a partir da formulação da teoria da pulsão de morte. Essa teoria, ao propor uma relação intrínseca entre vida e morte, encontra ecos nas concepções de autores como Bichat, Bernard e Canguilhem.

Por outro lado, Bocca (2020) discorda dessa interpretação, argumentando que Freud manteve, ao longo de sua obra, uma visão materialista e mecanicista fundamentada nos princípios da termodinâmica, como o princípio de inércia. Para Bocca, a noção de um vitalismo negativo deturpa a compreensão do pensamento freudiano, dado que os conceitos de inércia e espontaneidade vital são contraditórios.

Já Birman (1999) oferece uma leitura intermediária, propondo que, na década de 1920, Freud realiza a transição de uma perspectiva vitalista para uma perspectiva “mortalista”, na qual vida e morte se entrelaçam. Segundo Birman, a relação com o Outro seria o fundamento da pulsão de vida, contraposta à tendência ao inorgânico. Para ele, “a vida seria da ordem da transmissão e não uma qualidade inerente ao organismo humano” (p. 24), o que sugere que, na ausência do Outro, o organismo estaria mais próximo da morte do que da vida.

Portanto, o pensamento freudiano parece comportar múltiplas interpretações no que diz respeito às tradições mecanicistas e vitalistas aplicadas aos fenômenos mentais.

O presente artigo tem como objetivo investigar a questão: *a inteligência artificial (IA) pode substituir o psicanalista?* Essa reflexão será conduzida

A inteligência artificial pode substituir o psicanalista?

considerando os critérios que fariam a metapsicologia oferecer margem ou não para a redução do sujeito a um mecanismo.

A crescente presença de instrumentos técnicos na saúde mental frequentemente resulta na redução dos aspectos psicossociais e subjetivos em favor de modelos computacionais que priorizam variáveis quantificáveis, negligenciando as experiências pessoais de sofrimento (Bezerra Júnior, 2020). Embora as contribuições tecnológicas sejam inegáveis, há um risco de excessos, nos quais os pacientes se tornam “iPacientes” (Verghese, 2008), cuja atenção clínica se concentra na geração de índices quantificáveis em detrimento da exploração do sofrimento subjetivo e interpessoal.

No contexto psicanalítico, o desafio contemporâneo ultrapassa o conceito de “iPaciente” e chega ao de “iPsicanalista”, ou seja, à possibilidade de um chatbot desempenhar a função analítica. Diante da proliferação de tratamentos psicoterapêuticos mediados por IA que reivindicam referência à psicanálise (Capoulade & Pereira, 2020), a psicanálise é convocada a responder se um robô pode assumir a função de psicanalista e, em caso negativo, a justificar essa impossibilidade. Essa análise deverá destacar os limites da visão computacional do sujeito e sua aplicabilidade no tratamento psíquico, reafirmando a centralidade da experiência subjetiva na psicanálise.

Do que trata o computacionalismo?

No campo da filosofia da mente, diversos esforços têm sido realizados para caracterizar as condições que garantem a identidade de tipo dos estados psicológicos (Block & Fodor, 1980). A tese da identidade defende que não há diferença ontológica fundamental entre estados mentais e estados físicos do cérebro, baseando-se no axioma de que propriedades mentais (por exemplo, sentir dor) são isonômicas às propriedades físicas de primeira ordem (por exemplo, o disparo de fibras c). Caso essa tese seja verdadeira, ela sugere a dependência da mente à integridade do mundo físico, condicionando a explicação dos fenômenos mentais aos mesmos princípios que regem o funcionamento do universo material. Essa visão, portanto, tem profundas implicações sobre a maneira como compreendemos a natureza da mente e do mundo (Armstrong, 2002).

Nesse contexto, variantes do fisicalismo e do behaviorismo filosófico postulam, respectivamente, que dois organismos podem estar em estados psicológicos análogos apenas se alguns de seus estados físicos ou disposições comportamentais forem idênticos (Block & Fodor, 1980). Contudo, tal perspectiva ignora a possibilidade de estados psicológicos serem compartilhados entre organismos mesmo na ausência de equivalência em seus estados materiais, sejam eles biológicos ou comportamentais. Em outras palavras, não há impedimento teórico para que organismos morfologicamente distintos ou em estados físicos diversos em determinado momento apresentem o mesmo estado psicológico.

É plausível, por exemplo, que a experiência da dor seja compartilhada por espécies biologicamente distintas, como répteis, aves e mamíferos.

Na filosofia analítica, é amplamente aceito que as propriedades mentais têm realizabilidade múltipla. A tese da realizabilidade múltipla afirma que um único tipo mental pode ser realizado por diversos tipos físicos distintos. Essa ideia se fundamenta na “extrema probabilidade” (Putnam, 1979) de variações nas correlações entre propriedades psicológicas e suas realizações físicas em diferentes formas de vida. Um exemplo recorrente é o caso da dor. Muitos filósofos sustentam que uma ampla variedade de propriedades físicas, estados ou eventos, que não compartilham características comuns no nível físico, podem realizar a mesma experiência de dor. Essa tese desempenhou um papel crucial nos argumentos mais influentes contra as primeiras teorias que identificavam estados mentais com estados cerebrais (teorias psiconeurais ou de identidade mente-cérebro). Simultaneamente, forneceu suporte inicial aos argumentos favoráveis ao funcionalismo (Putnam, 1979).

Baseada nas noções de “multiplicidade” (isto é, variabilidade de propriedades) e “realizabilidade”, essa tese tornou-se central nos debates sobre se a mente pode ser reduzida ao cérebro, assim como na discussão sobre a possibilidade de redução de fenômenos tratados por disciplinas especializadas aos paradigmas das ciências físicas. A hipótese da realizabilidade múltipla é, portanto, um paradigma central nas discussões sobre a identidade de propriedades e os limites da redução de fenômenos por meio de disciplinas que operam com paradigmas de domínio estreito, propriedades específicas ou amplamente disjuntivas.

Embora modelos experimentais em pesquisas com animais sugiram a viabilidade de translação com base no parentesco filogenético entre espécies, não está claro se as mesmas regiões neuroanatômicas responsáveis pela dor ou pelo medo em ratos, como a amígdala e a ínsula, desempenham funções equivalentes em humanos e outros primatas (Server et al., 2018). Ao considerar estados comportamentais, as diferenças tornam-se ainda mais evidentes, uma vez que humanos podem derivar prazer da dor (Freud, 1924/2011) e responder a estímulos aversivos com um repertório muito mais amplo e inespecífico do que ratos em experimentos controlados.

Dessa forma, surge a questão: como é possível que a agregação de milhões de neurônios, individualmente insensíveis, dê origem à consciência subjetiva? Essa questão, amplamente debatida por neurocientistas e filósofos, reforça a ideia, cada vez mais disseminada, de que o cérebro seria a “base causal” dos fenômenos mentais. No entanto, até o presente, o que se tem são amplas correlações entre a atividade cerebral e a consciência. Assim, se “correlação não é o mesmo que causalidade” (Fuchs, 2018, p. 73), também não se pode assumir que a coocorrência, mais ou menos simultânea, de

A inteligência artificial pode substituir o psicanalista?

evidências materiais e psicológicas implique, necessariamente, uma relação causal direta (Fodor, 1983).

Dessa perspectiva, dois fenômenos podem manter uma relação de concomitância dependente, como Freud (1891/2013, 1915/2010b) sugeriu em relação aos processos psíquicos e fisiológicos, sem que, necessariamente, um “produza” o outro. Isso desafia as tentativas das neurociências de demonstrar uma relação causal direta entre cérebro e consciência. Mas, afinal, que tipo de estado corresponde à sensação subjetiva da dor?

Putnam (1979) propõe que a resposta à afirmação “A dor é um estado cerebral” deve permitir enunciados do tipo “dor é A”, onde “dor” e “A” não sejam sinônimos, e verificar se é possível encontrar uma formulação aceitável em termos empíricos e metodológicos. Ele rejeita que a dor seja um estado físico-químico do cérebro, ou mesmo de todo o sistema nervoso, sustentando que ela constitui um estado distinto, pois, sob a perspectiva funcionalista, o conhecimento de estados de um sistema não depende de suas realizações físicas. Nesse sentido, um sistema é uma organização funcional com entradas, saídas e estados internos bem definidos (Block & Fodor, 1980). A dor, então, é um estado funcional que emerge em sistemas semelhantes a autômatos probabilísticos, quando estímulos sensoriais adequados são processados (Putnam, 1979), analogamente ao funcionamento de um termostato frente a variações térmicas (Deacon, 2012). Para o funcionalismo, não importa o desconhecimento dos processos internos, desde que o comportamento do sistema seja controlável via estímulos e respostas. Assim, a dor é um estado funcional, e os organismos que a sentem são comparáveis a máquinas de Turing ou computadores (Putnam, 1979).

A filosofia da mente e da psicologia foi influenciada por três variantes do funcionalismo. A análise funcional utiliza uma estratégia explicativa que decompõe sistemas em partes para descrever fenômenos com base nas capacidades e na integração dessas partes. O funcionalismo de representação computacional compara processos mentais ao funcionamento de computadores, entendendo estados psicológicos como representações sistemáticas do mundo, similares a programas de computador. Já o funcionalismo metafísico focaliza a natureza da mente, abordando questões fundamentais, como “O que é dor?”, e compartilhando preocupações com o behaviorismo e o fisicalismo (Block & Fodor, 1980).

Se o computacionalismo consiste na tese de que os estados mentais são formas de processamento informacional logicamente estruturadas, ele poderia ser classificado, mutuamente, como uma versão modificada do funcionalismo (Chalmers, 1996) ou como integrante das teorias da cognição cujo programa, relacionado à psicologia, distingue-se do funcionalismo metafísico (Fodor, 1983; Piccinini, 2009). Entre os pressupostos mais difundidos e fundamentais

da ciência cognitiva está o de que a mente humana (ou a composição mente-cérebro) constitui algum tipo de mecanismo: um dispositivo físico composto por subsistemas funcionalmente especificáveis.

Essa proposta remonta à análise da modularidade cognitiva, introduzida por Jerry Fodor em *The modularity of mind* (1983), que institui a decomposição funcional – isto é, a análise do sistema global a partir de suas partes funcionalmente especificáveis – como projeto central para uma ciência da mente. Essa ciência deduz a psicologia humana do ponto de contato entre as diversas teorias sobre a arquitetura e o funcionamento cognitivo da espécie. Como a cognição é considerada modular quando, e somente quando, específica a um domínio, a tese da modularidade admite graus variados, uma vez que as funções mentais podem aparecer mais ou menos agregadas, como no caso da atenção, que frequentemente está acompanhada pela consciência, mas nem sempre pela memória (Samuels, 2006).

Aliada a características como o processamento de informações por meio de representações simbólicas (isto é, letras e números) e as prescrições operantes sobre a produção de resultados cognitivos (por exemplo, algoritmos), tal perspectiva propõe uma explicação abrangente dos fenômenos mentais por intermédio de modelos computacionais. Essa hipótese de trabalho, que estabelece uma isonomia entre o cérebro (e a mente nele subsumida) e o computador, parte da premissa de que a performance nas tarefas desempenhadas por sistemas biológicos constitui princípio heurístico suficiente para a generalização das capacidades explicativas extraídas do desempenho de sistemas artificiais (Piccinini, 2009).

O computacionalismo, contudo, parece ser, em si, produto de uma dupla redução: uma redução psicofísica, que subordina os estados subjetivos da mente aos estados materiais do cérebro; e outra, funcionalista, que submete tais estados aos algoritmos computacionais operados por representações simbólicas.

A crítica de Deacon à teoria computacional para o problema da mente

É verdade que “as ferramentas de pedra e os símbolos devem ser os dois arquitetos da transição *Australopithecus-Homo*, e não as suas consequências” (Deacon, 1997, p. 348), fazendo de nossa espécie, por origem, uma espécie-ciborgue, semelhante aos castores, na medida em que suas barragens transformaram seus antepassados terrestres em animais semiaquáticos. Contudo, ao longo da evolução no uso de ferramentas, é inegável que, desde a década de 1990, quando o computador da IBM Deep Blue venceu uma partida de xadrez contra o campeão mundial Garry Kasparov, temos observado uma

A inteligência artificial pode substituir o psicanalista?

crescente interface em nossas relações com dispositivos eletrônicos e informacionais (Deacon, 2012).

No limite, a capacidade da máquina de vencer, exclusivamente por meio de algoritmos, o campeão mundial de xadrez não apenas trouxe a percepção de que os computadores são mais rápidos do que os seres humanos na manipulação de signos, mas também gerou a intuição equivocada de que a suposta cognição da máquina – se é que ela existe – poderia ser superior à humana de maneira geral. Embora Kasparov tenha conseguido, em algumas ocasiões, empatar o jogo, seu esforço foi considerável, “enquanto a máquina não sentia nada” (p. 100). Fatos como esses, somados aos desenvolvimentos subsequentes de outras formas de inteligência artificial, impulsionaram o crescimento das abordagens computacionais para o problema da mente.

A computação é um processo em que uma parte do mecanismo tem o efeito de alterar outra (Deacon, 2012), a qual, por sua vez, altera outra, com o intuito de obter uma “instrução”. Nesse contexto, a computação opera de acordo com algoritmos, que são sequências finitas de ações executáveis, definidas em uma determinada ordem, capazes de, por meio de símbolos, ábacos ou computadores, produzir algo significativo ou representativo para o seu “usuário”. Assim como um software, o algoritmo é independente de sua incorporação física particular, funcionando apenas como um padrão físico. Nesse sentido, as instruções fornecidas por um algoritmo são descrições do que deve ocorrer; são “abstrações” de alguma forma, e não propriamente um processo físico em si mesmas (Deacon, 2012). Essa concepção sobre os algoritmos, que forneceram as bases necessárias para o surgimento da era computacional, foi a mesma que levou alguns cognitivistas a nutrirem a esperança de que os processos computacionais responderiam à lacuna explicativa sobre as relações entre mente e corpo.

Nessa concepção reducionista, a compreensão da mente humana não seria substancialmente diferente da de um zumbi (Chalmers, 1996; Deacon, 2012). Nossa experiência em primeira pessoa e nosso senso de escolha seriam uma mera ilusão, epifenômicos em relação a processos equivalentes a algoritmos. Toda ação particular seria uma “inferência dedutiva” (Deacon, 2012, p. 90) de sequências finitas, predefinidas e necessárias. Trata-se da visão de um pré-formacionismo universal das leis fundamentais da física e da química, que relegaria os fenômenos subjetivos a nenhum papel causal no universo. Nesse caso, se não há subjetividade no sentido mais básico que essa palavra pode designar, muito menos haveria eficácia simbólica.

Isso porque, nas variantes fiscalista e funcionalista do reducionismo, a lógica e a matemática seriam as estruturas fundamentais da existência, as quais edificam de maneira antecipada absolutamente todos os desdobramentos possíveis, tornando os processos relacionados à vida e à mente plenamente

previsíveis a partir de instâncias passadas. Tal como no demônio de Laplace, a criação do universo teria inscrita em suas próprias leis a impossibilidade de criação do *novo*. Por essa razão, embora uma concepção da vida e da mente como análogas à computação tenha proporcionado grandes avanços no nosso conhecimento sobre processos complexos, envolvendo uma vasta gama de variáveis e combinações, ela mesma pouco esclarece as questões relativas à capacidade dos selves de expressarem variabilidade normativa, uma precondição inclusive do processo evolutivo (Deacon, 2012).

Da mesma forma que o desenho de cada uma das letras e palavras com as quais escrevemos – isto é, itens físicos que independem de uma incorporação física particular, mas que são arranjados numa certa ordem de acordo com os interesses vivos de seus usuários –, os softwares computacionais só existem a partir de um fim externo que os representa e os organiza. Igualmente, o potencial para que essa manipulação de símbolos se torne propriamente uma informação só existe a partir do fato de que esse potencial é interpretado por alguém, ou seja, pelo seu usuário. Quem é esse “alguém” que interpreta o computador? Se nós, a vida e a mente, somos também apenas um padrão algorítmico, como afirmam algumas teorias computacionais, deveríamos concluir que um algoritmo é capaz de interpretar outro algoritmo?

Deacon (2012) critica as teorias computacionais da mente, inspirando-se, entre outras influências, na proposta do filósofo John Searle, particularmente em seu experimento heurístico denominado Chinese Room. Nesse experimento, Searle critica as teorias computacionais da mente por meio do seguinte cenário:

Somos observadores de uma sala onde há uma pessoa que não conhece o idioma chinês. Nossa tarefa é avaliar se há, em uma terceira sala (o quarto chinês), alguém com quem a pessoa observada está se comunicando. A pessoa na sala se comunica por meio de uma fenda na parede que separa as duas salas, trocando folhas de papel com caracteres chineses. Essa pessoa compara os símbolos nas folhas com um manual, que orienta quais símbolos devem ser respondidos, mas não fornece seu significado. A partir das instruções do manual, ela devolve uma resposta escrita em chinês pela fenda.

Nós, como observadores, devemos avaliar se há alguém no quarto chinês que entende o significado dos símbolos chineses. Podemos fazer essa avaliação? Ou estamos fadados a supor que no quarto chinês haja uma pessoa ou um robô que também compara os símbolos com um manual e responde cegamente, sem compreender seu significado? Dada essa configuração, é impossível saber se há alguém no quarto chinês compreendendo o significado das informações ou apenas um computador que manipula os signos por meio de isomorfismos, como a pessoa observada faz.

A inteligência artificial pode substituir o psicanalista?

Se seguíssemos os parâmetros do Teste de Turing – que define a inteligência como a capacidade de manipular símbolos de maneira indistinguível da de um humano –, concluiríamos que há uma inteligência no quarto chinês, seja ela proveniente de uma máquina ou de uma pessoa. No entanto, seríamos incapazes de avaliar se alguém realmente entende chinês.

Ao compreendermos que é diferente um sujeito entender chinês e apenas comparar entradas de ideogramas para gerar saídas, percebemos que a mera estrutura formal da língua não é capaz de capturar o componente de sentido de uma expressão comunicativa. Essa reflexão revela que a sintaxe não é condição suficiente para gerar semântica. A crítica ao computacionalismo demonstra que, embora uma tecnologia possa parecer inteligente, isso não a torna uma agência individuada, capaz de sentir afetos e responder a eles de maneira normativa, como nosso psiquismo faz.

De fato, ao buscarmos distinguir entre a mera transmissão de informações e a atuação de um self interpretante que confere significado a essas informações, o desfecho do processo revela-se menos importante do que a maneira como ele se desenrola. A questão que se coloca deve ser abordada a partir do ponto de vista genético (Freud, 1939/2018): como se origina essa diferença essencial entre a emergência dos processos de subjetivação e o funcionamento mecânico?

Para que uma máquina desempenhe sua função, como observa Deacon (2012), é crucial que seus componentes sejam altamente específicos e organizados de maneira igualmente específica. Quanto menos restritas forem as possibilidades de operação de cada parte, maior será o risco de perda da funcionalidade. Analogamente, a integridade e as funções de um ser vivo não são garantidas por adições ao que poderia acontecer, mas por *constraints* (restrições) aos estados potenciais. Se desaparefamos uma máquina, ela tende a perder sua função; da mesma forma, se as células replicam seu material genético fora dos padrões restritos, podem surgir tumores potencialmente letais.

Tanto em mecanismos quanto na vida, encontramos uma definição negativa das funções, ou seja, a necessidade de *constraints* frente ao fluxo constante de degradação dos materiais, conforme a lei da entropia. No entanto, pouco entendemos sobre o processo que estabelece essas restrições. Essa ideia é contundentemente expressa por Canguilhem, que afirma: “Quanto mais comparamos os seres vivos com máquinas automáticas, parece que melhor compreendemos a função, embora compreendamos menos a gênese” (2012, p. 128).

De acordo com Deacon (2012), existe uma diferença fundamental entre processos orgânicos e mecânicos na formação das partes e do todo, e consequentemente na gênese das restrições do sistema. Inspirado por Kant, Deacon argumenta que os organismos têm suas partes constituídas de maneira

recíproca, onde cada parte é simultaneamente o fim e o meio das demais. A constituição de uma parte do organismo ocorre por meio de uma sinergia sistêmica, na qual os componentes são produzidos mutuamente ao longo do tempo, sendo seu agrupamento em um todo organizado uma representação a priori, que tem a capacidade de restringir essas partes para manter seu funcionamento dentro de certos parâmetros.

De maneira oposta, segundo a lógica da engenharia, as partes são inicialmente independentes e se tornam interdependentes apenas quando organizadas em um todo, o que não faz delas simultaneamente causa e efeito umas das outras. Além disso, assim como as letras e palavras que utilizamos para escrever – itens físicos dispostos em uma ordem que reflete os interesses dos usuários –, os softwares computacionais existem apenas em função de um propósito externo que os representa. Como afirma Deacon, “a aparente agência do computador é efetivamente apenas a agência deslocada de algum projetista humano” (2012, p. 101), de modo que o mestre enxadrista não jogava contra um computador, mas contra um exército de programadores de jogos de xadrez. O que representa as relações entre as partes de um mecanismo de IA não é o todo que elas formam, mas o design feito por um projetista, que deve minimizar extrinsecamente a quantidade de ruídos que podem afetar a operação do sistema físico em busca de uma função específica.

Ao contrário de um sistema mecânico, as partes de um organismo “estão em fluxo constante, sendo constantemente sintetizadas, danificadas e substituídas, enquanto o todo persiste” (Deacon, 2012, p. 169). Em máquinas, montamos extrinsecamente componentes com materiais e formas cada vez mais resistentes ao fluxo de transformação, visando à sua durabilidade. Os materiais das máquinas são selecionados para resistir às flutuações termodinâmicas e evitar a deterioração. Em contraste, os seres vivos, embora dependam de condições termodinâmicas restritas para sua persistência, devem permanecer abertos a esses fluxos para autorregenerar as *constraints* que impedem sua dissolução, processo esse que é intrinsecamente representado. A transformação é, portanto, uma condição intrínseca à sua conservação. Cangulhem expressa bem essa diferença ao afirmar: “Em um organismo, observamos ... fenômenos de autoconstrução, autoconservação, autorregulação e autorreparação. No caso da máquina, a construção lhe é estranha e supõe a engenhosidade do mecânico” (2012, p. 125).

Tomemos como exemplo os neurônios. Esse tipo de célula evoluiu gradualmente a partir de células menos diferenciadas, tanto na filogênese quanto na embriogênese/ontogênese. Por razões como essa, os neurônios são transmissores de informação “desgovernados, ruidosos e modestamente confiáveis” (Deacon, 2012, p. 499). Com frequência, eles introduzem ruídos no input original, tanto pela falta de um design pré-formado que se alinhe com

A inteligência artificial pode substituir o psicanalista?

suas propriedades de sinalização quanto pela necessidade de realizar funções metabólicas simultaneamente à transdução neuroquímica da informação. Dado que contamos com um processador de informações composto por bilhões de nós de rede pouco confiáveis, é improvável que a coerência global do sistema nervoso e sua relação com o corpo e o ambiente se baseiem em um processo similar ao da computação. Podemos questionar: se os neurônios funcionassem como os componentes de um computador, mas com transdutores de informação tão imprecisos, isso não indicaria que os seres humanos seriam absolutamente estúpidos? Se quisermos evitar essas conclusões (nem tão) disparatadas, devemos investigar por que, apesar de serem individualmente pouco confiáveis e manipularem inputs de maneira diferente dos computadores, os neurônios, quando integrados em um cérebro, conseguem produzir outputs minimamente confiáveis.

Mecanismo não é motor

Os fenômenos distintivos da vida e da mente não podem ser compreendidos por divisões estanques entre partes e todo; eles demandam uma abordagem interdependente e processual. Diferentemente da engenharia, na qual as partes são previamente montadas para constituir um todo a posteriori, no caso do self as partes emergem apenas após a formação de um todo indiferenciado, que se diferencia ao longo do tempo. Um exemplo disso é a embriogênese: inicialmente composta por material inespecífico, a estrutura embrionária se segmenta por representações intrínsecas do sistema e gera células diferenciadas. Por outro lado, na lógica mecânica, a organização e a função das partes são determinadas por *constraints* extrínsecas, vinculadas a um trabalho ou representação alheios às partes, fruto do projeto de um engenheiro. O psiquismo, uma vez que está necessariamente inserido no processo vital, parece transcender essa determinação extrínseca. Existe uma intencionalidade intrínseca ao ser vivo, que cria as *constraints* necessárias à sua variabilidade, fecundidade e manutenção – sua normatividade vital, conforme Canguilhem (2012), ou sua vida erótica regulada pelo princípio do prazer, segundo Freud (1920/2010a).

Como vimos, para Ruz (2019), Bocca (2020) e Birman (1999), a ideia freudiana de que a tendência originária do vivente é a morte desafia a proposição central do vitalismo, que sustenta a primazia da tendência vital sobre a morte e o retorno aos mecanismos inanimados. No entanto, a solução apresentada por Birman (1999) – a de que a dependência do sujeito em relação ao Outro evita o retorno ao inorgânico – parece insuficiente. A questão que surge é: quem, inicialmente, transmitiu esses dons ao Outro? Não estaríamos, assim, diante de uma regressão infinita (Zaidhaft & Ortega, 2023)? Ou o Outro

original seria, então, algo além da matéria – um “mecânico do universo”? Ao mesmo tempo, se a subjetividade é vista como um algoritmo e o algoritmo como uma sequência predefinida e executável, surge a pergunta: “Quem ligou os botões?” (Deacon, 2012, p. 101).

Assim, não basta postular um “Outro” original que erogeniza; é crucial compreender as condições que permitem fundamentalmente a uma tendência de retorno ao inanimado engendrar, ainda que temporária e localizadamente, uma pulsão que se opõe a esse retorno. É necessário adotar uma perspectiva emergentista que reconheça tanto a continuidade quanto a descontinuidade entre a inércia mecânica e o aspecto vital da subjetividade, o que, na perspectiva de Deacon, assim como na de Freud (1920/2010a), implica um retorno à questão fundamental sobre a origem da vida. Em suma, a abordagem emergentista de Deacon evita a ideia de um “Outro” original, excepcional em suas origens em relação ao universo material e mecânico – um grande projetista que inscreveria a vida no terreno da morte –, sem, no entanto, tratar o sujeito como um epifenômeno de causas físico-químicas imediatas. A subjetividade ancora-se na causalidade mecânica, mas é um processo emergente, com propriedades distintas dessa base causal. Sem essa compreensão, corremos o risco de cair em um materialismo mecanicista, que elimina a subjetividade, ou em um vitalismo subjetivista, que assume dogmática e antropocentricamente um “Outro” com origens, propósitos e atributos excepcionais em relação ao universo mecânico (Zaidhaft & Ortega, 2023).

Além disso, ao considerar a crítica de Deacon (2012), é possível entender que a função do analista deve, para não ser substituível por robôs, focalizar a semântica, e não principalmente a sintaxe e demais elementos algoritmicamente formalizáveis. Embora algumas abordagens psicanalíticas priorizem a palavra e a gramática, tratando a língua como um sistema fechado, a verdade é que as máquinas manipulam significantes com um efeito pragmático semelhante ao dos humanos, mas sem significado interno ou compartilhado. A ausência de intencionalidade, normatividade ou senciência nas séries de signos manipuladas por robôs é evidente. As máquinas podem demonstrar inteligência, mas são absolutamente insensíveis – como pode um robô jogador de xadrez empatar com um mero humano e não sentir nada?

Portanto, a psicanálise seria inviável por meio de mecanismos de IA, desde que sua prática esteja fundamentada nos aspectos semântico, intencional e normativo das representações mentais, em contraste com abordagens que priorizam analisar o caráter sintático da associação livre e dos processos inconscientes. A psicanálise, portanto, deve concentrar-se no que as máquinas, por mais inteligentes que sejam, não têm: desejo, sentido e afeto.

A inteligência artificial pode substituir o psicanalista?

¿Puede la inteligencia artificial sustituir al psicoanalista? La crítica de Terrence Deacon a la teoría computacional de la mente

Resumen: La inteligencia artificial (IA) generativa, con su capacidad para realizar tareas léxicas antes restringidas al comportamiento humano, replantea el debate sobre la capacidad de los ordenadores para replicar el pensamiento, suscitando cuestiones para el psicoanálisis, tanto teóricas como clínicas. Los autores examinan las posibilidades y limitaciones de una práctica analítica mediada por algoritmos, planteándose la pregunta: ¿puede la IA sustituir al psicoanalista? Argumentan, de forma crítica con la teoría computacional para el problema de la mente, que el psicoanálisis sería inviable mediante mecanismos de IA, siempre y cuando su práctica se fundamente en los aspectos semánticos, intencionales y normativos de las representaciones mentales, en contraste con los enfoques que priorizan el análisis del carácter sintáctico de la asociación libre y los procesos inconscientes.

Palabras clave: Sigmund Freud, tecnologías digitales, computacionalismo, emergentismo, Terrence Deacon

Can artificial intelligence replace the psychoanalyst? Terrence Deacon's critique of the computational theory of the mind

Abstract: Generative artificial intelligence (AI), with its ability to perform lexical tasks previously restricted to human behavior, reframes the debate about the ability of computers to replicate thought, posing questions for psychoanalysis, both theoretical and clinical. The authors examine the possibilities and limitations of an analytical practice mediated by algorithms, considering the question: can AI replace the psychoanalyst? They argue, critically of computational theory for the problem of the mind, that psychoanalysis would be unfeasible by means of AI mechanisms, as long as its practice is grounded in the semantic, intentional and normative aspects of mental representations, in contrast to approaches that prioritize analyzing the syntactic character of free association and unconscious processes.

Keywords: Sigmund Freud, digital technologies, computationalism, emergentism, Terrence Deacon

L'intelligence artificielle peut-elle remplacer le psychanalyste ? La critique de la théorie computationnelle de la psyché par Terrence Deacon

Résumé : L'intelligence artificielle (IA) générative, avec sa capacité à effectuer des tâches lexicales auparavant limitées au comportement humain, recadre le débat sur la capacité des ordinateurs à reproduire la pensée, en posant des questions pour la psychanalyse, à la fois théoriques et cliniques. Les auteurs examinent les

possibilidades e as limites d'uma prática analítica mediada por algoritmos, em se posando a questão seguinte : a IA pode substituir o psicanalista ? Eles sustentam, criticando a teoria computacional para o problema da psique, que a psicanálise seria impossível por meio de mecanismos de IA, tanto que sua prática é baseada nos aspectos semânticos, intencionais e normativos das representações mentais, ao contrário das abordagens que dão prioridade à análise do caráter sintático da associação livre e dos processos inconscientes.

Mots-clés : Sigmund Freud, tecnologias digitais, computacionalismo, emergentismo, Terrence Deacon

Referências

- Armstrong, D. M. (2002). *A materialist theory of the mind*. Routledge.
- Bezerra Júnior, B. (2020). Tecnologias digitais, subjetividade e psicopatologia: possíveis impactos da pandemia. *Revista Latinoamericana de Psicopatologia Fundamental*, 23(3), 495-508.
- Birman, J. (1999). A dádiva e o Outro: sobre o conceito de desamparo no discurso freudiano. *Physis: Revista de Saúde Coletiva*, 9(2), 9-30.
- Block, N. & Fodor, J. A. (1980). What psychological states are not. In J. J. Katz, D. T. Langendoen & G. A. Miller (Orgs.), *The language of thought* (pp. 237-250). Harvard University.
- Bocca, F. V. (2020). Freud, um materialista mecanicista às últimas consequências. *Revista Latinoamericana de Psicopatologia Fundamental*, 23(2), 372-359.
- Canguilhem, G. (2012). *O conhecimento da vida* (V. L. Avellar, Trad.). Forense Universitária.
- Capoulade, F. & Pereira, M. E. C. (2020). Desafios colocados para a clínica psicanalítica (e seu futuro) no contexto da pandemia de covid-19: reflexões a partir de uma experiência clínica. *Revista Latinoamericana de Psicopatologia Fundamental*, 23(3), 534-548.
- Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. Oxford University.
- Deacon, T. (1997). *The symbolic species: the co-evolution of language and the brain*. W. W. Norton & Company.
- Deacon, T. (2012). *Incomplete nature: how mind emerged from matter*. W. W. Norton & Company.
- Fodor, J. A. (1983). *The modularity of mind*. MIT.
- Freud, S. (2010a). Além do princípio do prazer. In S. Freud, *Obras completas* (P. C. Souza, Trad., Vol. 14, pp. 161-239). Companhia das Letras. (Trabalho original publicado em 1920)
- Freud, S. (2010b). O inconsciente. In S. Freud, *Obras completas* (P. C. Souza, Trad., Vol. 12, pp. 99-150). Companhia das Letras. (Trabalho original publicado em 1915)
- Freud, S. (2011). O problema econômico do masoquismo. In S. Freud, *Obras completas* (P. C. Souza, Trad., Vol. 16, pp. 184-202). Companhia das Letras. (Trabalho original publicado em 1924)

A inteligência artificial pode substituir o psicanalista?

- Freud, S. (2013). *Sobre a concepção das afasias: um estudo crítico* (E. B. Rossi, Trad.). Autêntica. (Trabalho original publicado em 1891)
- Freud, S. (2018). Moisés e o monoteísmo. In S. Freud, *Obras completas* (P. C. Souza, Trad., Vol. 19, pp. 13-188). Companhia das Letras. (Trabalho original publicado em 1939)
- Fuchs, T. (2018). *Ecology of the brain: the phenomenology and biology of the embodied mind*. Oxford University.
- Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass*, 4(3), 515-532.
- Putnam, H. (1979). The mental life of some machines. In H. Putnam (Org.), *Mind, language and reality* (Vol. 2, pp. 408-428). Cambridge University.
- Ruz, F. H. (2019). Freud, la pulsión y lo viviente: en búsqueda de un vitalismo negativo. *Revista Latinoamericana de Psicopatología Fundamental*, 22(1), 95-116.
- Samuels, R. (2006). Is the human mind massively modular? In R. J. Stainton (Org.), *Contemporary debates in cognitive science* (pp. 37-56). Blackwell.
- Server, A., Reina, M. A., Boezaart, A. P., Prats-Galino, A., Coelho, M. E. & Sala-Blanch, X. (2018). Microanatomical nerve architecture of 6 mammalian species: is trans-species translational anatomic extrapolation valid? *Regional Anesthesia & Pain Medicine*, 43(5), 496-501.
- Souza Leite, P. C. B. (2022). Hiperconectividade e exaustão. *Jornal de Psicanálise*, 55(102), 127-147.
- Turkle, S. (1988). Artificial intelligence and psychoanalysis: a new alliance. *Daedalus*, 117(1), 241-268.
- Verghese, A. (2008). Culture shock: patient as icon, icon as patient. *New England Journal of Medicine*, 359(26), 2748-2751.
- Zaidhaft, E. & Ortega, F. (2023). Espíritos, afetos e biologia: o cuidado em saúde mental à luz do emergentismo de Terrence Deacon. *Revista Latinoamericana de Psicopatología Fundamental*, 26, e210436.

Recebido em 23/11/2024, aceito em 9/12/2024

Eduardo Zaidhaft
eduardozaidhaft@gmail.com

Ramon Reis
ramonrsferreira@gmail.com

Monah Winograd
monahwinograd@icloud.com

DOI: 10.69904/0486-641X.v58n4.08