UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE FILOSOFIA

YASMIN DOS SANTOS OLIVEIRA

Entre Simulação e Compreensão: As Implicações da Crítica de John Searle à Inteligência Artificial Forte

Uberlândia
2025

YASMIN DOS SANTOS OLIVEIRA

Entre Simulação e Compreensão: As Implicações da Crítica de John Searle à Inteligência Artificial Forte

Trabalho de Conclusão de Curso apresentado ao Instituto de Filosofia da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de licenciatura e bacharelado em Filosofia.

Área de concentração: Filosofia da Mente

Orientador: Leonardo Ferreira Almada

Uberlândia

2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

O48
2025

Oliveira, Yasmin dos Santos, 2001-
Entre Simulação e Compreensão: As Implicações da Crítica de John Searle à Inteligência Artificial Forte [recurso eletrônico] / Yasmin dos Santos Oliveira. - 2025.

Orientador: Leonardo Ferreira Almada.
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Uberlândia, Graduação em Filosofia.
Modo de acesso: Internet.
Inclui bibliografia.

1. Filosofia. I. Almada, Leonardo Ferreira,1981-, (Orient.). II. Universidade Federal de Uberlândia. Graduação em Filosofia. III. Título.

CDU: 1

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

YASMIN DOS SANTOS OLIVEIRA

Entre Simulação e Compreensão: As Implicações da Crítica de John Searle à Inteligência Artificial Forte

Trabalho de Conclusão de Curso apresentado ao Instituto de Filosofia da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de licenciatura e bacharelado em Filosofia.

Área de concentração: Filosofia da Mente

Uberlândia, 2025

Banca Examinadora:

_____

Prof. Dr. Leonardo Ferreira Almada **(**IFILO, Presidente**)**

_____

Prof. Dr. Fábio Coelho da Silva (IFILO)

_____

Professor Mestrando Luciano Henrique Moreira Santos (PGPSI-UFU)

Este trabalho é dedicado à mulher que sempre acreditou em mim. Reconheço ser impossível resumir tamanha grandeza numa palavra tão pequena: mãe.

AGRADECIMENTOS

Agradeço a oportunidade de ter acesso à educação superior pública e de qualidade, graças à contribuição da população e ao empenho dos professores que compõem o Instituto de Filosofia. Todo o corpo docente foi de extrema importância para que eu chegasse aqui, em especial os professores **Alcino Eduardo Bonella —** pelo incentivo e direcionamento **— Rafael Cordeiro Silva —** pelas correções e conselhos **— Fábio Coelho da Silva —** pelas indicações de leitura e dicas acadêmicas **—** e **Leonardo Ferreira Almada —** pela orientação, suporte e motivação. Todos vocês fizeram os meus olhos brilharem e o meu coração bater mais depressa, ao perceber o quanto me identificava com o que diziam. Minha jornada acadêmica não seria a mesma sem vocês.

Minha mãe, **Josefina Maria dos Santos**, que me incentivou a cada passo e me encorajou a insistir mesmo diante dos obstáculos que enfrentei. Você foi crucial na formação do meu caráter e da minha forma de enxergar o mundo. Dentre os feitos dos quais me orgulho, a maioria foi possível graças a você, que me deu o sustento e apoiou minhas aspirações. Não poderia ter chegado aqui sem você. Sua figura me inspira todos os dias.

Apollo César Abdelnur Alves, obrigada por me aturar em meus momentos de estresse, chateação e fadiga, e me apoiar quando estive cabisbaixa e desorientada. Você inspirou e impulsionou grande parte da minha produtividade!

Minhas parcerias de curso, que têm minha gratidão pelos momentos de descontração, conversas e apoio em questões burocráticas. Estar com vocês na UFU foi a cereja do bolo na minha formação, não só acadêmica como também pessoal. **Barbara Leandra Porto Mota**, **Bárbara Rafaelle Carvalho Santos**, **Beatriz Gomes Favoretto**, **Gabriel Carvalho da Silva** e vários outros amigos com os quais tive a honra de compartilhar momentos durante as aulas e além delas. A interação com vocês enriqueceu minha história acadêmica e as memórias da minha passagem na graduação serão para sempre marcadas com a alegria das trocas que tivemos.

Agradeço, por fim, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concedida durante minha graduação, cujo fomento foi essencial para que eu pudesse me dedicar aos estudos e contribuir em minhas pesquisas. Direcionar meus esforços para analisar e elaborar sobre os temas que me são relevantes é de extremo proveito e possível, claro, graças a dedicação exclusiva proporcionada pela bolsa estudantil.

RESUMO

A presente monografia examina algumas implicações filosóficas da crítica de John Searle à concepção de inteligência artificial forte, especialmente no que diz respeito à possibilidade de que máquinas possam possuir compreensão genuína e consciência. O objetivo do trabalho é demonstrar que, apesar dos avanços técnicos no campo da inteligência artificial, os sistemas computacionais não possuem os elementos fundamentais que caracterizam a mente humana, como a intencionalidade intrínseca e a experiência subjetiva. Para isso, adota-se uma metodologia de cunho teórico e analítico, baseada em revisão bibliográfica de textos filosóficos e científicos, com foco nas obras de John Searle e em debates contemporâneos sobre filosofia da mente, ciência cognitiva e inteligência artificial. A argumentação central estrutura-se a partir do experimento do quarto chinês, proposto por Searle em 1980, que ilustra a diferença entre manipulação sintática de símbolos e compreensão semântica. O trabalho também analisa o conceito de intencionalidade e sua distinção entre formas intrínsecas e derivadas, evidenciando que os computadores operam apenas por meio de regras formais e não por meio de significados compreendidos. A crítica de Searle aos materialismos reducionistas é explorada em paralelo à discussão sobre redes neurais artificiais, incluindo os desafios enfrentados por esses sistemas devido à ausência de um *Background* experiencial e de um contexto cultural, social e pessoal, características essenciais da cognição humana. Casos empíricos em que sistemas de inteligência artificial falham em tarefas interpretativas são apresentados como exemplos que sustentam a tese de que a simulação não equivale à compreensão. Os resultados indicam que, mesmo com avanços tecnológicos, as IAs operam via padrões estatísticos e correlações de dados, sem acesso ao *Background* humano, conjunto de capacidades não-representacionais que fundamentam a compreensão contextual. Conclui-se que a consciência permanece uma propriedade emergente de sistemas biológicos, e que atribuir estatuto ontológico a máquinas constitui um erro categorial, com implicações éticas e epistemológicas relevantes. O estudo reforça a necessidade de abordagens interdisciplinares para discutir os limites da simulação computacional e valorizar a singularidade da cognição humana.

Palavras-chave: consciência; intencionalidade; inteligência artificial.

ABSTRACT

The present monograph examines some philosophical implications of John Searle's critique of the conception of strong artificial intelligence, especially regarding the possibility that machines could possess genuine understanding and consciousness. The objective of the work is to demonstrate that, despite technical advances in the field of artificial intelligence, computational systems do not possess the fundamental elements that characterize the human mind, such as intrinsic intentionality and subjective experience. For this, a theoretical and analytical methodology is adopted, based on a bibliographic review of philosophical and scientific texts, with a focus on the works of John Searle and on contemporary debates about philosophy of mind, cognitive science, and artificial intelligence. The central argument is structured from the Chinese Room experiment, proposed by Searle in 1980, which illustrates the difference between syntactic manipulation of symbols and semantic understanding. The work also analyzes the concept of intentionality and its distinction between intrinsic and derived forms, showing that computers operate only through formal rules and not through understood meanings. Searle's critique of reductionist materialisms is explored in parallel with the discussion about artificial neural networks, including the challenges faced by these systems due to the absence of an experiential background and of a cultural, social, and personal context—essential characteristics of human cognition. Empirical cases in which artificial intelligence systems fail in interpretative tasks are presented as examples that support the thesis that simulation is not equivalent to understanding. The results indicate that, despite technological advances, artificial intelligences operate through statistical patterns and data correlations, without access to the human Background, a set of non-representational capacities that underlie contextual understanding. It is concluded that consciousness remains an emergent property of biological systems, and that attributing ontological status to machines constitutes a category mistake, with significant ethical and epistemological implications. The study reinforces the need for interdisciplinary approaches to discuss the limits of computational simulation and to value the uniqueness of human cognition.

Keywords: consciousness; intentionality; artificial intelligence.

SUMÁRIO

INTRODUÇÃO

Em um contexto de avanço nas ciências cognitivas e no desenvolvimento inicial da computação, especialmente a partir da segunda metade do século XX, muitos pensadores começaram a formular teses sobre a possibilidade de a Inteligência Artificial[1] um dia alcançar compreensão e consciência semelhantes às dos seres humanos. É nesse cenário que o filósofo norte-americano John Searle apresenta suas posições, debruçando-se sobre questões tais como a natureza da consciência e o que poderia distingui-la de uma IA. Seu objetivo de envolver a IA é justamente o que fez com que fosse escolhido como autor basilar dessa monografia.

É importante ressaltar que a IA conhecida pelo autor em sua época era de um nível distinto do que concebemos atualmente, uma vez que os computadores[2] ainda eram máquinas rudimentares, limitados para processamento de informações por meio de algoritmos matemáticos e regras sintáticas, sem o desempenho ou sofisticação conhecidos hoje. Além disso, até mesmo a internet era um invento recente[3], pouco acessível e não popularizada. Por isso, devemos pensar seu raciocínio frente às discussões que o autor acompanhou sobre a máquina de Turing[4] (referenciada por Turing especialmente no artigo *Computing Machinery and Intelligence* [Computadores e Inteligência], de 1950) e os debates sobre computação simbólica.

As argumentações e objeções contidas no ensaio de Searle intitulado *Minds, Brains, and Programs* [Mentes, Cérebros e Programas], de 1980, serão a base para analisar as características que impedem uma IA de ser idêntica à mente humana, evitando afirmações equivocadas que menosprezam a particularidade desta. A crítica de Searle tem como alvo principal certas versões do funcionalismo, especialmente o computacionalismo, que propõem que os estados mentais podem ser definidos unicamente pelas relações causais entre entradas de estímulos, estados internos e saídas comportamentais. Dentro dessa perspectiva, os

---

[1] IA, doravante.

[2] O primeiro computador programável do mundo data de 1941, mas a popularização dos computadores pessoais ocorreu posteriormente, a partir da década de 80 (Oliveira, 2011, p. 18-20).

[3] A internet foi criada em 1958 para a guerra fria, aprimorada na década de 70 para conectar apenas institutos de pesquisa e universidades e, a partir de 1991, refinada para ser comercializada (Oliveira, 2011, p. 25).

[4] [...] É um dispositivo teórico conhecido como máquina abstrata universal, concebida pelo matemático britânico Alan Turing muitos anos antes de existirem os modernos computadores digitais, tais como os conhecemos hoje. Por isso, é um modelo abstrato (teórico) de um computador, que se restringe apenas aos aspectos lógico-conceituais de seu funcionamento (memória, estados e transições etc.) e não à sua implementação física. [...] a ideia central que Turing queria demonstrar e argumentar a favor era a de que uma Máquina de Turing poderia simular o mesmo desempenho da mente humana, em qualquer tipo de sua atividade e não apenas no sentido de calcular. (Canal, 2012, p. 28)

processos mentais seriam análogos à manipulação de símbolos em um computador digital — ou seja, à relação entre hardware e software. Searle rejeita essa tentativa de reduzir a cognição humana a esse tipo de estrutura formal e algorítmica, argumentando que tais sistemas carecem de intencionalidade intrínseca e consciência genuína.

A linha de raciocínio construída por Searle por meio do experimento do quarto chinês intriga seus opositores que defendem o funcionalismo e o computacionalismo: imaginar uma situação em que estamos confinados em um quarto, recebendo um papel com símbolos chineses desconhecidos, juntamente com instruções que nos permitem manipulá-los com base em regras predefinidas. Embora as respostas geradas possam parecer coesas para um observador externo, aquele que está dentro do quarto organiza as respostas mas não entende absolutamente nada do conteúdo: apenas executa manipulações sintáticas sem acesso à semântica.

Esse experimento demonstra, desde o início, que mesmo ao responder corretamente às perguntas, a pessoa no quarto não adquire compreensão real do significado dos símbolos, semanticamente. Da mesma forma, Searle argumenta que uma máquina devidamente programada pode gerar respostas aparentemente inteligentes, mas sem uma compreensão genuína do que está processando. Isso ilustra, segundo ele, a limitação fundamental do computacionalismo e do funcionalismo — a incapacidade de explicar a consciência e a intencionalidade apenas com base em manipulações formais de símbolos.

Com esse experimento, o autor impulsionou a reflexão e objeção de pensadores como Daniel Dennett e Paul Patricia Churchland, na busca pela explanação das características exclusivas da cognição humana, não só para fins ontológicos, mas também para fins epistemológicos e éticos. Estabelecer se a consciência existe, o que ela é, e verificar se possui características intrínsecas (como intencionalidade[5] e subjetividade) foram algumas das motivações dos pensadores estimulados por sua tese. O debate sobre o que distingue a mente

---

[5] A intencionalidade é um termo herdado especialmente do filósofo Franz Brentano, que lança sua explicação sobre o que posteriormente é denominado intencionalidade como sendo "uma propriedade geral dos fenômenos psíquicos, que distingue essa classe de fenômenos da classe de fenômenos físicos. [...] Outra propriedade, comum a todos os fenômenos psíquicos, é que eles só são percebidos na consciência interna, enquanto a única percepção possível dos fenômenos físicos é a externa [...]. [A percepção interna tem] evidência imediata e infalível que corresponde somente a ela entre todos os modos de conhecer objetos da experiência. Portanto, quando dizemos que os fenômenos psíquicos são apreendidos pela percepção interna, quer dizer que sua percepção tem uma evidência imediata" (Brentano, 1935, p. 85-86, tradução nossa).
Posteriormente, esse termo foi investigado por Edmund Husserl, para fundamentar a fenomenologia: "A intencionalidade é aquilo que caracteriza a consciência no sentido forte, e que justifica ao mesmo tempo designar todo o fluxo de vivido [evento psicológico qualquer] como fluxo de consciência e como unidade de uma única consciência." (Husserl, 2006, p. 190)

humana das máquinas permanece central em Filosofia e nas ciências cognitivas em geral, especialmente diante das tentativas de modelar a IA com base em processos cognitivos.

Em outras palavras, considerando as questões ontológicas, epistêmicas, éticas e sociais que envolvem a humanidade, este trabalho buscará interpretar o posicionamento de Searle e aplicá-lo para uma análise da IA, pois reconhecer suas limitações, já denunciadas pelo autor no século passado, é um excelente ponto de partida para nos posicionarmos criticamente frente aos avanços digitais com os quais nos deparamos e evitar equívocos, como o de acreditar que uma IA desenvolvida possa substituir o pensamento humano. A tese de Searle sobre o que é a consciência traz reflexões para reafirmarmos a plasticidade da consciência humana e suas propriedades singulares e alheias a sistemas sintéticos.

METODOLOGIA

Tendo em vista que meu intuito é o de analisar a Inteligência Artificial (IA) à luz do experimento do quarto chinês de John Searle, proponho, neste trabalho, explanar as limitações das máquinas em alcançar compreensão e consciência genuínas bem como as implicações associadas ao desenvolvimento de tecnologias avançadas, destacando a importância de reconhecer as particularidades do pensamento humano diante do progresso digital.

Os objetivos específicos de minha pesquisa envolvem o cotejo de diferentes abordagens para compreender a relação entre a IA e o pensamento humano. Em primeiro lugar, proponho investigar as principais argumentações de John Searle apresentadas em seu ensaio *Minds, Brains, and Programs* [Mentes, Cérebros e Programas] (1980), com ênfase no experimento do quarto chinês, a fim de compreender suas objeções às teorias funcionalistas e as limitações inerentes às máquinas em simular a cognição humana.

Com base na análise bibliográfica conduzida a partir de uma leitura filosófica, proponho-me refletir sobre a importância de delimitar e diferenciar o pensamento humano. O objetivo é evitar o equívoco de considerar a inteligência artificial como substituta da mente humana e, ao mesmo tempo, fomentar um posicionamento crítico diante do progresso digital contemporâneo.

Minha monografia, então, adota uma abordagem teórico-conceitual, tendo como atividades centrais a revisão, comparação e articulação crítica dos textos selecionados para os debates sobre a IA e suas implicações adjacentes. A parte central da monografia — dedicada à exposição e análise dos textos — constitui os resultados da investigação, revelando os principais argumentos sobre consciência, intencionalidade e as limitações da IA à luz da filosofia da mente. Já as implicações éticas, sociais e epistemológicas decorrentes desses resultados são discutidas ao final, com o intuito de fomentar uma reflexão crítica sobre os desafios que acompanham o avanço tecnológico atual.

Portanto, inicialmente, farei uma revisão bibliográfica, com ênfase nos textos primários de John Searle, como *Minds, Brains, and Programs* [Mentes, Cérebros e Programas] (1980) e *The Rediscovery of the Mind* [A redescoberta da mente] (1992). Essa etapa inclui ainda a busca por artigos acadêmicos e publicações complementares, com o objetivo de contextualizar as contribuições desse autor no âmbito da filosofia da mente. A revisão bibliográfica não apenas mapeia os argumentos centrais do pensador, mas também identifica críticas e diálogos relevantes com suas teorias, oferecendo um panorama abrangente sobre o estado atual das discussões.

Em seguida, minha pesquisa se dedica à articulação crítica dos conceitos analisados. Essa etapa envolve a comparação entre as perspectivas de Searle e opositores, com foco nos pontos de convergência e divergência entre suas abordagens. Enquanto Searle enfatiza as limitações das máquinas em alcançar compreensão genuína, utilizando o experimento mental do quarto chinês para desbancar teorias funcionalistas, é possível ampliar o debate ao explorar os riscos existenciais e éticos associados aos inventos recentes. A articulação busca compreender como essas teorias, mesmo partindo de pressupostos distintos, podem ser utilizadas para fundamentar uma reflexão mais ampla sobre os avanços tecnológicos e suas implicações para a humanidade.

Dessa forma, portanto, viso a integração dos resultados das etapas anteriores em uma análise crítica que destaca as subjacentes ao desenvolvimento da IA. Essa abordagem metodológica, ao combinar revisão de literatura e articulação de conceitos, permite construir um arcabouço teórico para avaliar os limites e as possibilidades da IA reforçando a necessidade de um posicionamento crítico diante dos avanços tecnológicos.

Além da reconstrução conceitual, busco explorar de modo breve as implicações filosóficas e éticas contemporâneas do avanço da inteligência artificial, considerando seus impactos sociais, cognitivos e culturais. Assim, os resultados deste trabalho não se limitam à exposição teórica, mas incluem uma reflexão crítica sobre os rumos que a sociedade toma diante do uso crescente de tecnologias baseadas em IA, com o intuito de fomentar o pensamento autônomo e responsável frente às transformações digitais em curso.

1 O EXPERIMENTO DO QUARTO CHINÊS E A CRÍTICA AOS MATERIALISMOS

1.1 Apresentação do capítulo

A discussão sobre a IA e a consciência não só me intriga como também inflama diversas discussões desde seu surgimento no século XX. Neste capítulo, explorei as contribuições de John Searle para esse debate, com ênfase em sua crítica à concepção materialista da mente e à ideia de que sistemas computacionais podem, por si só, ser dotados de consciência ou compreender verdadeiramente o significado das informações que processam. Para isso, me voltarei para três obras suas, a saber: (i) *Minds, Brains, and Programs* [Mentes, Cérebros e Programas] (1980); (ii) *Intentionality – An essay in the philosophy of mind*, [Intencionalidade – Um ensaio sobre filosofia da mente] (1983); e (iii) *The Rediscovery of the Mind* [A redescoberta da mente] (1992). Contemplo esses textos com o intuito de conhecer quais são os argumentos que levaram Searle a afirmar que a base biológica do ser humano é o fator necessário para emergência da consciência.

Inicialmente, apresentei os conceitos fundamentais empregados na argumentação, diferenciando a Inteligência Artificial Fraca da Inteligência Artificial Forte. A partir disso, a fim de contestar a concepção forte da IA, expus a lógica do experimento do quarto chinês, que busca demonstrar que um programa computacional pode manipular símbolos sem necessariamente compreendê-los. Esse exercício expõe a diferença entre a sintaxe, que um computador pode processar, e a semântica, que envolve a compreensão do significado.

Elucidei também as críticas de Searle às correntes materialistas da filosofia da mente, como o *behaviorismo*, funcionalismo e o eliminativismo, que sustentam diferentes formas de reducionismo mental[6]. Searle argumenta que essas teorias falham em explicar a consciência e sua relação intrínseca com o comportamento e o cérebro e desejei demonstrar aqui alguns dos motivos dos quais o autor se vale. Outro ponto crucial abordado é a relação consciência-comportamento-cérebro. A intencionalidade é o ponto-chave para compreender por que um fenômeno intrínseco da mente consciente não pode ser capturado meramente pela manipulação de símbolos. Essa perspectiva se opõe à visão de que o comportamento externo de um sistema poderia ser suficiente para inferir estados mentais reais.

---

[6] Tese segundo a qual os estados e processos mentais podem ser totalmente explicados ou substituídos por descrições em termos físicos, biológicos ou computacionais. Em outras palavras, visa a redução de fenômenos como a consciência, a intencionalidade ou os sentimentos a processos cerebrais, comportamentais ou algoritmos.

Por fim, explanarei o conceito de consciência e a subjetividade dos fenômenos mentais, enfatizando como a experiência subjetiva e a qualidade fenomenológica dos estados mentais são elementos que, defende Searle, são irredutíveis. Ao longo deste capítulo, então, buscarei oferecer explicação para os termos caros para a análise crítica da posição do autor, buscando conhecer suas afirmações e a articulação dos conceitos para criar condições de, no capítulo seguinte, avaliarmos alguns aspectos do cenário atual da tecnologia e refletir um pouco sobre o debate contemporâneo acerca da IA.

1.2 Os conceitos empregados por John Searle

Como os conceitos filosóficos são frequentemente polissêmicos e sujeitos a diferentes interpretações, é fundamental contextualizar e esclarecer os termos utilizados por Searle para compreender sua linha de raciocínio. Por isso, no decorrer deste texto, conceitos como **consciência**, **intencionalidade** e **inteligência artificial** serão explicados em detalhe, fornecendo a base necessária para a compreensão do tema. Vale ressaltar que o próprio autor se preocupa em definir previamente os termos que utiliza, evidenciando que esse esforço de esclarecimento não deve ser desprezado. Sua argumentação depende diretamente da compreensão assertiva dos conceitos empregados, garantindo que suas ideias sejam interpretadas de forma precisa ao seu propósito.

1.2.1 Inteligência Artificial Fraca e Forte

Primeiramente, e a fim de especificar o objeto de sua argumentação, Searle conceitua a IA em duas categorias: (i) fraca e (ii) forte. Para explicar o que considera ser uma IA fraca, o autor discorre:

> De acordo com a IA no sentido fraco, o principal valor do computador para o estudo da mente reside no fato de que este nos fornece uma ferramenta extremamente poderosa. Por exemplo, ele nos permite formular e testar hipóteses de maneira mais rigorosa e precisa do que antes. (Searle, 1996, p. 63)

Essa categorização ajuda a compreender que, por proporcionar acessibilidade e eficiência aprimoradas, esse tipo de IA apoia os avanços da humanidade na compreensão da mente, já que seu objetivo é desempenhar agilmente um escopo de funções limitadas e

específicas. São exemplos de IA no sentido fraco: programas que fornecem informações de clima atualizadas automaticamente, sites de buscas na internet com precisão e velocidade superiores aos de um humano, calculadoras etc. São consideradas fracas, acima de tudo, por não tentarem replicar o funcionamento complexo de uma mente, mas apenas realizar bem as funções pré-programadas, sendo, inclusive, incapazes de realizar outras tarefas para além do proposto. Não obstante, essa categoria não é explorada por Searle, pois sua preocupação não reside nesse tópico.

A preocupação e o objeto de investigação de Searle é a IA no sentido forte, cuja pretensão é a de ser propriamente uma mente humana com estados mentais. Essa classificação de inteligência é altamente contestada pelo autor, que elucida sobre o alvo dos golpes de sua crítica no seguinte excerto:

> De acordo com a IA no sentido forte, o computador não é meramente um instrumento para o estudo da mente. Muito mais do que isso o computador adequadamente programado é uma mente, no sentido de que, se lhe são dados os programas corretos pode-se dizer que eles entendem e que eles têm outros estados cognitivos. Conforme a IA no sentido forte, uma vez que o computador programado tem estados cognitivos, os programas não são meros instrumentos que nos capacitam testar explicações psicológicas: os programas constituem as próprias explicações. (Searle, 1996, p. 64)

É então a partir dessa conceituação que toda a linha de raciocínio e, consequentemente, o objeto de estudo deste trabalho pretendem se construir. Isso porque, dada a complexidade e a dificuldade de categorização objetiva acerca do funcionamento da mente humana e da natureza do pensamento, a discussão sobre a comparação funcionalista e fisicalista deve ser analisada com mais escrutínio. Afinal, como sustentar que nossa mente pode ser replicada em toda sua complexidade e magnitude a partir de um arranjo específico de peças?

Para elucidar acerca da consciência, Searle interpreta a mente por meio de uma descrição biológica, conferindo-a como uma característica oriunda do processo evolutivo no seguinte trecho: "[...] sou um certo tipo de organismo com uma certa estrutura biológica (física e química) e esta estrutura, em termos causais, é capaz, sob certas condições, de produzir a percepção, a ação, a compreensão, o aprendizado e outros fenômenos intencionais." (Searle, 1996, p. 84-85). Essa estrutura mencionada pelo autor é o que possibilita a emergência da intencionalidade em que a consciência se direciona para as coisas externas e objetos.

Sua tese biológica de mente é ainda reafirmada no seguinte trecho: a "intencionalidade, é um fenômeno biológico o qual deve ser tão causalmente dependente da bioquímica específica de suas origens como o é a lactação, a fotossíntese ou quaisquer outros fenômenos biológicos."

(Searle, 1996, p. 92). Isso afirma que uma estrutura como o *hardware*, que concebemos hoje pela tecnologia, não é capaz de fazer com que o programa existente no computador seja um produto próprio, tal qual os estados mentais são produto próprio do cérebro. Adiante iremos nos dedicar a compreender mais alguns dos conceitos que levaram o autor a chegar nessa conclusão.

1.2.2 O experimento do quarto chinês

Como fio condutor para pensar o experimento proposto em seu artigo *Minds, Brains, and Programs* [Mentes, Cérebros e Programas], Searle cita o programa (*software*) criado por Roger Schank na década de 70, que buscava simular o processo humano de compreensão de informações implícitas em histórias, como: 'onde está o humor na tirinha?'. Diante de seus opositores, para os quais a IA não só compreende a história como também é modelo para explicar o funcionamento do processo humano de entendimento de histórias, o autor postula que a compreensão não deve ser confundida com uma simulação de compreensão, sendo esta última o procedimento que o programa faz.

Para desbancar tais teorias, o experimento nos convida a imaginar-nos dentro de um quarto, em que se recebe um papel com símbolos chineses desconhecidos e outro com significações na língua nativa, que permitem manipular os símbolos e significados para produzir respostas que, para um observador externo, soam como inteligentes. Esse experimento nos mostra que, mesmo respondendo às perguntas, o indivíduo no quarto continua sem entender a língua chinesa, pois está operando como um programa, apenas decodificando e produzindo combinações, sem criar nada de novo. Analogamente, é assim que se encontra uma máquina, apenas produzindo conteúdos inteligentes sem os realmente compreender.

Isso demonstra ainda que não é possível utilizar a IA como modelo para entender o funcionamento da mente humana, uma vez que na relação com questões na sua língua nativa haveria uma instância de entendimento que não há na relação com questões em chinês (modelo pelo qual um computador atuaria). A defesa é a de que há um limite do que pode ser feito pelo computador, a saber, "operações computacionais especificadas sobre elementos puramente formais" (Searle, 1996, p. 69). Para além desse limite, está a compreensão, possível para os seres humanos, que podem realmente entender **intencionalmente** o que estão produzindo.

Um dos pontos centrais quando analisamos a teoria searleana consiste na asserção de que o computador possui apenas sintaxe, mas não possui compreensão semântica. Por exemplo: "se você digita: 2+2 igual?, ele vai apresentar '4'. Mas ele não tem ideia que '4' significa 4, ou

que isto signifique alguma coisa.” (Searle, 1996, p. 90). Isso quer dizer que se enxergamos alguma intencionalidade na IA, na verdade é porque está apenas na cognição daqueles que lhes construíram ou lhe conferem metaforicamente essa propriedade. Em outras palavras: simular compreensão não equivale a ter intencionalidade e doravante irei expor mais alguns conceitos que reafirmam essa posição.

1.2.3 Correntes materialistas e as objeções de Searle

Searle publica, em 1992, a obra *The Rediscovery of the Mind* [A redescoberta da mente], que busca aprofundar ainda mais sua empreitada para elucidar a consciência, esclarecendo aí a distinção entre as características intrínsecas de um objeto (nesse caso, a consciência), que independem do observador, e aquelas que lhe são extrínsecas e, portanto, existem somente enquanto objeto de relação com um observador externo.

Pode-se dizer, *grosso modo*, que essa é a primeira parte da linha de raciocínio, que objetiva defender e descartar modos de pensarmos o que é a consciência. Searle evidencia sua percepção de que há uma crise na tradição filosófica, e a qual insiste em pesquisar a mente humana de forma objetiva e considerando apenas a observação em terceira pessoa.

As correntes de pensamento que foram influenciadas pelo positivismo lógico — movimento filosófico do início do século XX, liderado pelo Círculo de Viena, que defendia que a filosofia deveria adotar uma abordagem científica rigorosa e eliminar especulações metafísicas — passaram a estudar a mente exclusivamente sob uma perspectiva externa e empírica. Segundo Searle, esse enfoque comete o erro de ignorar a dimensão subjetiva da experiência consciente, reduzindo a mente a comportamentos observáveis ou a meras descrições físico-químicas. Destacam-se aqui as correntes materialistas, que defendem que tudo o que existe é, em última instância, material, e as behavioristas, que sustentam que apenas os comportamentos observáveis podem ser objeto legítimo de estudo científico da mente, rejeitando explicações mentalistas e introspectivas.

Contrapondo-se a essas abordagens, Searle sustenta que a mente não pode ser plenamente compreendida sem considerar os fenômenos internos e a perspectiva da primeira pessoa, aspectos essenciais da consciência e da intencionalidade.

Alguns de seus opositores materialistas, sobre os quais incide a crítica searleana, pertencem a diferentes correntes filosóficas dentro do materialismo. Entre eles, encontram-se os funcionalistas, como Daniel Dennett, Jerry Fodor e Hilary Putnam, que defendem que os

estados mentais podem ser explicados em termos de funções computacionais e relações causais. Há também os behavioristas, como B.F. Skinner, que rejeitam a introspecção e tratam os estados mentais como meras descrições do comportamento observável. Por fim, os materialistas eliminativistas, como Paul e Patricia Churchland, argumentam que conceitos como crença e desejo são apenas construções linguísticas equivocadas e que, no futuro, serão substituídos por explicações neurocientíficas mais precisas. Searle critica todas essas abordagens por, segundo ele, ignorar ou subestimar a realidade da consciência e da intencionalidade.

O cerne de *The Rediscovery of the Mind* [A redescoberta da mente] é a preocupação em desmistificar a confusão entre o método epistemológico para compreender a mente e a ontologia da mente, que pertencem a categorias distintas, uma vez que a ontologia da mente pode ser verificada apenas na relação de primeira pessoa. Para tratar desse assunto, como é de se esperar, o dualismo de propriedades[7] e o dualismo de substâncias[8] são termos mencionados, uma vez que a aversão à herança cartesiana é um elemento para entender o sucesso das teorias materialistas. Isso pois, visando a se livrar da ideia de haver uma instância metafísica da mente, o que não lhes era desejável, excluem-na totalmente.

De outro modo, relações causais e comportamentos não estão necessariamente ligados. Um exemplo que ilustra essa tese é a de que podemos estar sentindo uma dor e não demonstrar nenhuma reação para não sermos percebidos ou ainda um ator que é capaz de simular uma dor que não está sentindo durante a interpretação de um papel. Sobre esse ponto, Searle acrescenta: "crenças e desejos algumas vezes causam ações, mas não há nenhuma ligação essencial. A maioria das crenças e desejos nunca resulta em ações" (Searle, 2006, p. 89).

> Qual é, exatamente, a importância do comportamento para o conceito de mente? **Ontologicamente falando, comportamento, papel funcional e relações causais são irrelevantes para a existência de fenômenos mentais conscientes. Epistemicamente**, de fato aprendemos acerca dos estados mentais conscientes de outras pessoas, e o fazemos **em parte**, a partir de seu comportamento. **Causalmente**, a consciência serve para mediar as relações causais entre os estímulos de *input* e o comportamento *output*; [...] (Searle, 2006, p. 103, grifos do autor).

[7] "[...] o dualismo de propriedades diz que há dois tipos essencialmente diferentes de *propriedades* no mundo. [...] No caso da mente, o dualismo de propriedade é defendido por aqueles que argumentam que a natureza qualitativa da consciência não é meramente outra maneira de categorizar estados do cérebro ou do comportamento, mas um fenômeno genuinamente emergente." (Robinson, 2023, tradução nossa).

[8] No contexto da mente, um dualista de substância acredita que a mente não é apenas um conjunto de pensamentos ou estados mentais, mas é "uma substância imaterial acima e além de seus estados imateriais" (Robinson, 2023, tradução nossa) — um 'eu' distinto do corpo físico, que produz os pensamentos. Assim, este dualismo afirma não apenas que os estados mentais são diferentes dos estados físicos, mas que pertencem a uma substância imaterial separada.

Em outras palavras, para Searle, é possível ter fenômenos mentais conscientes sem demonstrar comportamento e é possível demonstrar determinado comportamento sem necessariamente ter o fenômeno mental consciente.

Searle empenhou-se em explicar o significado do materialismo: "[...] querem negar a existência de quaisquer fenômenos mentais irredutíveis no mundo. [...] negar que haja quaisquer propriedades fenomenológicas irredutíveis, tais como consciência ou ***qualia*** (Searle, 2006, p. 44)."

Ainda segundo o autor (Searle, 2006, p. 45), se aceitamos ser um fato da física que o mundo "é constituído de partículas físicas em campos de força", deveríamos também aceitar que "somos todos conscientes e que nossos estados conscientes têm propriedades fenomenológicas irredutíveis bastante específicas". Isso porque as mencionadas partículas físicas em campos de força não são visíveis para nós e, mesmo assim, as temos como um fato.

Searle busca legitimar os fenômenos mentais da mesma forma que legitimamos os fenômenos digestivos do estômago para ir contra a ideia de que os termos como **subjetividade, consciência e intencionalidade** são inconsistentes e ineficientes para tratar objetivamente a mente (Searle, 2006, p. 45). Ou seja, as entidades que não se encaixam perfeitamente em explicações baseadas apenas em leis físicas ou computacionais (como crenças, desejos, emoções etc), são legitimadas da mesma forma que legitimamos os fenômenos digestivos do estômago, por exemplo.

Em outras palavras, assim como aceitamos que o estômago possui um vocabulário próprio para descrever suas funções digestivas (deglutição, peristaltismo etc), Searle afirma que deveríamos reconhecer que os fenômenos mentais possuem um vocabulário específico legítimo, defendendo que a validade de um fenômeno não depende exclusivamente da sua explicação em termos estritamente físicos ou computacionais. Sua defesa aparece também no seguinte trecho: "[...] A neurobiologia da consciência deve provavelmente se revelar pelo menos tão restrita quanto, digamos, a bioquímica da digestão" (Searle, 2006, p. 136).

Assim, o autor prescreve que os fenômenos mentais necessitam uma terminologia própria, uma vez que a mente é um fenômeno biológico e "nossos estados conscientes têm propriedades fenomenológicas irredutíveis bastante específicas" (Searle, 2006, p. 45). Isso implica que a tentativa de evitar termos mentalísticos que atestam implicitamente a existência da consciência é desaprovada por Searle, pois considera perda de tempo buscar um vocabulário alternativo para fenômenos que são propriedades intrínsecas ao mental (como dor, fome, frio, nojo etc.).

Ou seja, a empreitada materialista de substituir "Eu sinto fome" por "Meu sistema nervoso liberou grelina[9]" não resolve a questão de que experienciamos um fenômeno específico com propriedades fenomenológicas irredutíveis que nos faz sentir vontade de ingerir comida. Searle postula que "Deveríamos simplesmente admitir, em primeiro lugar, os fenômenos mentais (e portanto físicos), da mesma maneira como admitimos os fenômenos digestivos no estômago" (Searle, 2006, p. 46). Se aceitamos que o vocabulário do estômago tem conceitos e fenômenos próprios do digestivo, deveríamos aceitar que o vocabulário próprio da mente, com termos como dor, crença, intenção, vontade, também descreve uma realidade legítima, com um modo de existência que não pode ser completamente traduzido para os termos da neurobiologia ou da física.

Temos ainda a afirmação de que:

> A consciência é uma propriedade emergente, ou de nível superior, do cérebro, no sentido absolutamente inócuo 'de nível superior' ou 'emergente', no qual a solidez é uma propriedade emergente de nível superior de moléculas de H_2O quando estas estão em uma estrutura cristalina (gelo), ao passo que a liquidez é, de forma semelhante, uma propriedade emergente de nível superior de moléculas de H_2O quando estas estão, *grosso modo,* girando em torno umas das outras (água). A consciência é uma propriedade mental e, portanto, física, do cérebro, no sentido em que a liquidez é uma propriedade de sistemas de moléculas" (Searle, 2006, p. 25-26).

Isso nos mostra que, embora recuse abordagens reducionistas que eliminem a dimensão fenomenológica da experiência consciente, Searle afirma que a consciência é uma propriedade física do cérebro. Aqui vale ressaltar que o autor postula que, no tocante a consciência, está "falando de sua **irredutibilidade de acordo com modelos padrões de redução**. [...] [sem] deixar de considerar, *a priori*, a possibilidade de uma revolução intelectual importante [...] segundo a qual a consciência seria redutível" (Searle, 2006, p 179, grifos do autor).

Sua asserção é que

> [...] causas neurofisiológicas de tipo idêntico teriam efeitos mentalísticos de tipo idêntico. Assim, tomando o famoso exemplo do cérebro-dentro-da-cuba, se você tivesse dois cérebros que fossem tipo-idênticos até a última molécula, então a base causal do mental garantiria que tivessem os mesmos fenômenos mentais. [...] a superveniência do mental em relação ao físico é determinada pelo fato de que estados físicos são causalmente suficientes, embora não necessariamente causalmente necessários, para os estados mentais correspondentes. Esta é apenas outra forma de afirmar que, no que diz respeito

[9] "[...] Essa substância, produzida pelos neurônios e principalmente pelo estômago, foi revelada por cientistas britânicos como o fator que estimula o apetite" (Gardenal, 2002, p. 3).

> a esta definição de superveniência, identidade de neurofisiologia garante identidade de mentalidade; contudo, identidade de mentalidade não garante identidade de neurofisiologia. (Searle, 2006, p. 180).

Isso quer dizer que o estado físico "basta" para trazer o estado mental à existência. No entanto, o autor reconhece a possibilidade de que um determinado estado mental possa ser causado por diferentes configurações físicas. Uma configuração física específica sempre causará o mesmo estado mental, mas o mesmo estado mental poderia talvez ser causado por configurações físicas diferentes. Esta é uma forma de expressar o endosso à ideia de realizabilidade múltipla, onde um estado de nível superior (o mental) pode ser implementado por diferentes estados de nível inferior (o físico/neurofisiológico).

Para ilustrar a irredutibilidade da consciência, Searle aponta (2006, p. 164) que diferentemente do calor, que passou por um processo de redefinição conceitual ao ser compreendido como o movimento médio das moléculas, eliminando de sua definição os aspectos subjetivos da sensação térmica, a consciência resiste a esse tipo de redução. No caso do calor, a ciência conseguiu distinguir claramente entre o fenômeno físico objetivo e a aparência subjetiva que ele causa, tratando esta última como um efeito colateral irrelevante para a descrição da realidade física. Isso permitiu uma redução ontológica eficaz, pois a sensação foi deixada de lado e o 'calor real' passou a ser entendido apenas em termos de energia cinética. Essa abordagem funciona bem porque a aparência subjetiva de calor não é essencial para o funcionamento ou a definição do fenômeno físico que a causa.

No entanto, como argumenta Searle, embora possamos aceitar que a consciência tenha causas físicas, sua natureza fenomenológica não pode ser capturada por descrições objetivas de terceira pessoa como fazemos com o calor. Assim, a irredutibilidade da consciência não se mostra um mistério metafísico, mas uma consequência da própria estrutura dos nossos conceitos e métodos de definição científica. Capturar a consciência pelo nosso paradigma científico seria como descrever um som apenas pelas ondas sonoras, sem jamais tocar na experiência de ouvi-lo. A empreitada falha pois a dor não é apenas causada por padrões específicos de atividade cerebral; ela é uma experiência subjetiva.

Desta forma, implica-se que "a ontologia efetiva dos estados mentais é uma **ontologia de primeira pessoa**" (Searle, 2006, p. 28, grifo nosso). No entanto, na prática, a aplicação do ponto de vista de terceira pessoa (característica marcante de nosso paradigma científico, conforme mencionado), torna desafiador diferenciar algo que realmente possui uma mente, como um ser humano, de algo que apenas se comporta como se tivesse uma, como um computador (Searle, 2006, p. 28-29). Entretanto, existem objeções a essa explicação.

Com o intuito de golpeá-las posteriormente, algumas correntes opositoras da fenomenologia, anteriormente mencionadas neste texto de modo breve, são discorridas, analisadas e categorizadas por Searle. A primeira dessas é o (i) *behaviorismo* lógico (ou comportamentalismo), o qual "sustenta que a mente é somente comportamento e disposições para comportamento [...], consiste na concepção de que afirmações sobre fenômenos mentais podem ser traduzidas em afirmações sobre comportamento possível e real" (Searle, 2006, p.52). Para o autor (2006, p. 54), "o absurdo do *behaviorismo* repousa no fato de que ele nega a existência de quaisquer estados mentais internos além do comportamento exterior". O *behaviorismo* lógico deixa de lado a sensação subjetiva e qualitativa dos fenômenos mentais e, portanto, apresenta lacunas explicativas.

Outra teoria contra a qual Searle se posiciona é o (ii) funcionalismo, que, em sua versão computacionalista, apresenta a mente como um programa de computador. Essa vertente sustenta que os estados mentais são definidos por suas funções causais — ou seja, pelas relações entre entradas, estados internos e saídas — independentemente do substrato físico que os realiza. Assim, a mente poderia, teoricamente, ser implementada em qualquer sistema computacional adequado, sem necessidade de um cérebro biológico. Searle critica essa perspectiva ao argumentar que os processos mentais não podem ser reduzidos meramente a manipulações formais de símbolos, pois a consciência e a intencionalidade são fenômenos que emergem de forma específica em sistemas biológicos, estando profundamente enraizados na estrutura do cérebro humano. Portanto, para ele, não é possível que esses fenômenos surjam em qualquer outro substrato de maneira equivalente.

Um dos autores de destaque da corrente funcionalista é Jerry Fodor, que se empenhou em explicitar sua defesa ao funcionalismo no artigo *The Mind-Body Problem* [O Problema Mente-corpo], de 1981. Fodor salienta que o funcionalismo é uma terceira via, alternativa às posições dualistas e materialistas, pois concebe os estados mentais em termos de suas funções causais e computacionais, independentemente da substância física que os realiza. Diferentemente do dualismo, que postula a mente como uma entidade separada da matéria, e do materialismo reducionista, que equipara estados mentais a estados físicos específicos do cérebro, o funcionalismo permite que diferentes sistemas físicos possam ter estados mentais equivalentes, desde que desempenhem as mesmas funções no processamento de informações.

Fodor resolve a questão sobre máquinas não possuírem propriedade de compreensão semântica admitindo que, na verdade, os símbolos são dotados de propriedades semânticas. Isso significa dizer que os símbolos processados por qualquer sistema (seja o cérebro humano, ou uma máquina de Turing) são intrinsecamente semânticos, de modo que, para reconhecer uma

crença em um sistema, basta que este manipule o símbolo de acordo com o seu significado. Em outras palavras: se dizemos que ‘o céu é azul’, isso basta para provar que essa crença tem seu significado ligado ao estado mental que a origina. Desta forma, manipular/processar o símbolo é equivalente a representar os estados mentais contidos no sistema. Isso iguala máquinas a humanos no que diz respeito a possuir uma mente.

No entanto, Fodor reconhece que o funcionalismo não consegue rebater de modo eficiente as objeções elaboradas sob o que se denomina ‘viés do enigma do espectro invertido’ pois:

> [...] é possível imaginar dois observadores semelhantes em todos os aspectos psicológicos relevantes, com a única exceção de que as experiências subjetivas que possuem o conteúdo qualitativo do vermelho para um observador teriam o conteúdo qualitativo do verde para outro observador. O comportamento de ambos não revela a diferença porque todos os dois veem um tomate maduro e um pôr do sol flamejante como sendo de cor semelhante e chamam essa cor de ‘vermelho’ (Fodor, 2011, p. 14).

Com isso, percebemos que o funcionalismo se mostra limitado ao lidar com a dimensão subjetiva e qualitativa da experiência consciente, tal como exemplificado no problema do espectro invertido. Esse experimento mental evidencia que dois indivíduos poderiam ter funções mentais idênticas — responder da mesma forma a estímulos visuais, por exemplo — mas experienciar cores de maneiras radicalmente distintas, sem que isso fosse detectável externamente. Isso revela a incapacidade do funcionalismo de explicar os *qualia*, ou seja, os aspectos fenomenológicos da experiência — como é para alguém ver o vermelho, sentir dor ou ouvir uma melodia. A teoria, portanto, negligencia elementos essenciais da consciência, justamente aqueles que constituem sua natureza mais íntima e experiencial.

Conforme mencionado, a crítica ao materialismo e ao behaviorismo se faz presente em todo o discurso dessa obra. Isso pode ser percebido em excertos como “É um erro supor que sabemos da existência dos fenômenos mentais em outras pessoas somente pela observação do seu comportamento” (Searle, 2006, p. 35) e “Comportamento ou relações causais para comportamento não são fundamentais para a existência de fenômenos mentais” (Searle, 2006, p. 38). Dessa forma, é possível verificar que podemos sentir determinada emoção ou crença e não demonstrar para observadores externos, e demonstrar, por outro lado, um comportamento sem ser verdadeiramente afetado por alguma emoção ou crença.

A terceira das mais proeminentes correntes é o (iii) Materialismo eliminativo (ou eliminacionista), o qual nega a existência da mente e, consequentemente, invalida a psicologia popular, ou seja, a maneira pela qual explicamos os estados mentais e o comportamento humano

intuitiva e cotidianamente, implicitamente atestando a existência de emoções e crenças, por exemplo. Em outras palavras, por meio da linguagem do senso comum, descrevemos nossas "entidades mentais ordinárias" (Searle, 2006, p. 70): nossas aspirações, sentimentos, intenções etc. Entretanto, conforme supracitado, o materialismo eliminativo advoga que uma ciência cognitiva avançada será capaz de nos fazer desacreditar nessas entidades, dando vazão a uma explicação estritamente neurocientífica.

O autor dedica um apêndice para escrutinar esses ataques à psicologia popular, que nutrem a afirmação de que estamos equivocados quando fazemos afirmações sobre nossos conteúdos mentais, pois não são objetivas o suficiente para fazer disto uma ciência. Essa corrente de pensamento anti-psicologia-popular conta com apoiadores como Paul Feyerabend e Paul M. Churchland. Dentre as proposições contra a psicologia popular, a afirmação de que os objetos de estudo como crenças e desejos não existem traz inquietação não só para Searle mas também para mim.

Essa posição é problemática para Searle sobretudo por ignorar a realidade fenomenológica da experiência subjetiva e a intencionalidade da mente. Afinal, se crenças e desejos não existem, como podemos explicar a estrutura da ação humana ou a comunicação cotidiana? Aceitar a eliminação desses conceitos parece contraintuitivo, pois implicaria uma revisão drástica da forma como compreendemos a cognição, a linguagem e até mesmo a moralidade.

1.2.4 Relações entre consciência, comportamento e cérebro

Para desafiar a ideia de que o mundo pode ser todo explicado objetivamente pela ciência e mostrar que existem instâncias que não podemos reduzir ao domínio positivista, como nossas experiências subjetivas, o autor refuta de forma contundente os argumentos prescritos pelos materialistas, e o faz provocando a reflexão por meio do **experimento mental do cérebro de silício**. O experimento consiste em verificar as consequências e possibilidades diante de uma situação em que os neurônios de um cérebro em degeneração fossem progressivamente substituídos por circuitos feitos de silício.

Pensando no cenário do cérebro de silício, Searle apresenta três possibilidades distintas ao discutir a relação entre **processos cerebrais**, **estados mentais** e **comportamento observável**. De certo modo, há aqui uma interpretação e conexão dos aspectos considerados relevantes pelas correntes mencionadas (*behaviorismo*, funcionalismo e materialismo). A primeira possibilidade seria a do indivíduo (nomearei de José) que teve parte do cérebro

substituída por silício e manteve todas as suas capacidades mentais, como pensar, experienciar e lembrar, sem qualquer alteração na vida mental. Isso significaria que os circuitos artificiais teriam as mesmas capacidades causais que o cérebro biológico, possibilitando tanto a manutenção dos estados mentais quanto a reprodução do comportamento externo.

A segunda possibilidade considera o cenário no qual a experiência consciente de José é reduzida, mas sem que isso afete o comportamento observável. Nesse caso, os circuitos substituiriam apenas parte das funções causais do cérebro, resultando em um organismo que continua a agir normalmente, mas sem vivenciar estados mentais da mesma forma. Isso implicaria que apenas o comportamento externo seria preservado, sem a reprodução completa da consciência. E, por fim, a terceira possibilidade sugere o oposto: a experiência consciente permaneceria intacta, mas o comportamento externo seria comprometido devido a uma paralisia. Aqui, os circuitos artificiais preservariam os estados mentais, mas falhariam em reproduzir a capacidade motora, impedindo a manifestação física do comportamento. Conforme o autor: "A finalidade dessas três variações do experimento de pensamento é ilustrar as **relações causais entre processos cerebrais, processos mentais e comportamento externamente observável**". (Searle, 2006, p. 102, grifo nosso). Dito de outro modo, esse experimento evidencia que a cognição não pode ser reduzida apenas à execução de funções externas, pois a consciência dá sua contribuição nesse processo.

Um outro apontamento interessante feito por Searle consiste em afirmar que nem todos os fatos são empiricamente observáveis, como: em primeira pessoa, não podemos saber como é poder usar a ecolocalização, como faz um golfinho, por exemplo. Em outras palavras, há uma subjetividade intrínseca ao observador em primeira pessoa, que não pode ser acessada em terceira pessoa, analisando apenas seu comportamento.

Uma questão que muito me incomodou ao começar a estudar a teoria searleana foi: como podemos afirmar, filosoficamente, que uma máquina não tem consciência quando não conseguimos experenciar em primeira pessoa o que é ser uma máquina? Encontra-se um acalento para essa inquietação no seguinte trecho:

> Estou bastante seguro de que a mesa à minha frente, o computador que uso diariamente, a caneta-tinteiro com que escrevo e o gravador para o qual dito são completamente inconscientes, mas, logicamente, não posso *provar* que são inconscientes, e nem pode fazê-lo nenhuma outra pessoa. (Searle, 2006, p. 115).

Frente a esse desafio, se me impôs outra questão: como podemos conferir que existe cognição em outros seres quando só podemos consultar a ontologia da mente acessando nossa

própria experiência subjetiva? Searle (2006) sugere que essa suposição pode ser sustentada por um fundamento empírico, o que significa que a justificativa não vem de um raciocínio puramente filosófico ou metafísico, mas de observações do mundo real. Em outras palavras, baseamo-nos no comportamento observável e em evidências científicas, como as semelhanças neurológicas, para inferir que outras mentes existem. Aqui, encontramos uma das pretensões de Searle: "[...] dar uma explicação do fundamento 'empírico' que temos para supor que outras pessoas e animais superiores tenham fenômenos mentais conscientes mais ou menos como os nossos próprios" (Searle, 2006, p. 107)

A resolução para o problema das outras mentes é possível na medida em que nos relacionamos com outros seres e percebemos que suas reações e comportamentos diante de determinados estímulos são semelhantes às reações que temos quando experienciamos tal ocorrido. Por exemplo, ao sentir um beliscão em nossa pele e sentir dor, a reação tende ser a de escaparmos de quem nos infringiu o desconforto. Algo semelhante ocorre, também, quando um cachorro é beliscado, o que nos faz intuir que ele sente algo parecido ao que sentimos quando somos beliscados: dor.

Há uma distinção interessante feita por Searle que apoia a compreensão do que realmente deve ser entendido como intencionalidade da consciência e o que é mera metáfora sem sentido literal. Eis a divisão em três categorias: a primeira delas é a (i) **intencionalidade intrínseca**, que diz respeito ao fenômeno que enunciamos para descrever um conteúdo mental intencional, como por exemplo "eu sinto sede". Essa afirmação representa um conteúdo intencional real e a utilizamos de forma intuitiva cotidianamente.

No entanto, (ii) uma vez que somos dotados de intencionalidade, atribuímos, por analogia, intencionalidade às coisas, mesmo que não acreditemos que a possuem literalmente. Isso faz com que não seja estranho dizer "essa planta está com sede", para expressar que, se estivéssemos na situação da planta, estaríamos com sede. Nós estendemos nossa intencionalidade para os instrumentos com os quais lidamos. Assim, estamos diante do que Searle chama de **intencionalidade *como se,*** pois é *como se* a planta tivesse intencionalidade. Também o fazemos ao mencionar que um programa 'reconhece uma imagem' ou 'entende uma pergunta', pois estamos alegando que é *como se* ele reconhecesse e entendesse. A seguir, isso é esclarecido nas palavras do autor:

> [...] Intencionalidade intrínseca é um fenômeno que seres humanos e determinados outros animais têm como parte de sua natureza biológica. Não é uma questão de como são tratados, ou como se concebem a si mesmos, ou de que forma preferem descrever a si mesmos. É simplesmente um fato evidente em tais animais que, por exemplo, algumas vezes fiquem com **sede**

> ou **fome**, **vejam** coisas, **temam** coisas etc. Todas as expressões em itálico nas frases anteriores são empregadas para indicar estados intencionais intrínsecos. É muito conveniente usar o jargão da intencionalidade para falar sobre sistemas que não a têm, mas que se comportam como se tivesse. [...] digo, sobre meu computador, que sua **memória** é maior do que a **memória** do computador que eu tinha no ano passado. Todas essas atribuições são perfeitamente inofensivas, e sem dúvida acabarão produzindo novos significados literais quando as metáforas se tornarem mortas. Mas é importante enfatizar que essas atribuições são psicologicamente irrelevantes, porque não implicam a presença de nenhum fenômeno mental. A intencionalidade exposta em todos esses casos é puramente **como-se.** (Searle, 2006, p. 118-119, grifos do autor)

Por sua vez, a (iii) **intencionalidade derivada** é aquela que literalmente tem intencionalidade atribuída, mas não é intrínseca ao sistema que a enuncia, como quando um falante nativo de português diz "Em inglês, '*I feel thirsty*' significa 'eu sinto sede'". Isso porque, "o significado linguístico é uma forma real de intencionalidade, mas não é intencionalidade intrínseca. É derivado da intencionalidade intrínseca daqueles que usam a língua" (Searle, 2006, p. 118). Ou seja, a intencionalidade derivada diz respeito à atribuição de intencionalidade a fenômenos não mentais, como textos que associamos ao significado, mapas que evocam a atitude de associação com o território ilustrado, retratos que evocam a atitude de associação com o indivíduo retratado. Também pode ser verificada quando visualizamos uma imagem com ilusão de ótica e prontamente dizemos 'A imagem está se mexendo!'. Em todos esses casos, a intencionalidade não surge nos objetos de forma intrínseca — ela é resultado da intencionalidade dos agentes humanos, relativa às capacidades do observador externo. Isso quer dizer que uma máquina possui informação no mesmo sentido que uma biblioteca ou um livro a tem, relativas à nossa capacidade de interpretação.

Searle assevera que "[...] o preço de rejeitar a distinção entre intencionalidade intrínseca e **como-se** é o absurdo, porque torna mental tudo no universo" (Searle, 2006, p. 122, grifos do autor). Isso significa que, se rejeitássemos haver intencionalidade real, consideraríamos apenas a intencionalidade *como-se*, atribuindo agência e intenção até mesmo a objetos inanimados. Para relatar um exemplo extremo desse equívoco, Searle menciona uma pedra. No entanto, para nosso contexto, prefiro mencionar um computador. O equívoco de recusar a singularidade da intencionalidade intrínseca pode nos levar a afirmar que um computador *gera* respostas *como-se* tivesse intencionalidade, quando, na realidade, a intencionalidade é derivada dos programadores e usuários humanos.

1.2.5 O conceito de consciência em Searle

Devido à polissemia atribuída ao termo consciência, que pode ter diversas definições como sendo a dimensão moral das nossas ações ou o estado de percepção de algo interno ou externo a nós, torna-se essencial esclarecer seu uso no contexto em questão. "Existem diferentes graus de consciência" (Searle, 2006, p. 124), uma vez que, se estamos em estado de vigília, isso difere do estado de sonolência, se estamos embriagados, isso difere da sobriedade.

Awareness[10] é um sinônimo aproximado para definir consciência, mas ainda assim não é fiel o suficiente pois (i) está mais ligado à cognição e conhecimento do que à noção de consciência é (Searle, 2006, p.124) e (ii) podemos estar cientes de algo de modo inconsciente[11] (Searle, 2006, p. 125). O autor assevera:

> Os estados conscientes sempre têm um conteúdo. Ninguém nunca pode ser somente consciente; ao contrário, quando alguém é consciente tem que haver uma resposta à pergunta: 'De que esse alguém é consciente?' Mas o 'de' de 'consciente de' nem sempre é o 'de' de intencionalidade. Se estou consciente de uma batida na porta, meu estado consciente é intencional, porque faz referência a algo além disto mesmo, a batida na porta. Se estou consciente de uma dor, a dor não é intencional, porque não representa nada além dela mesma (Searle, 2006, p.125).

A pretensão aqui é igualar o conceito de consciência aos conceitos científicos sobre a teoria atômica da matéria e a teoria da evolução biológica, que são notoriamente aceitos por convencer por meio das evidências que atestam.

Existem diferentes formas de explicarmos os fenômenos, como por exemplo: por que a água ferve? (a) porque coloquei a panela sob a chama; ou (b) porque "a energia cinética transmitida pela oxidação de hidrocarbonetos para as moléculas de H_2O fez com que estas se

---

[10] Conhecimento, no sentido de estar ciente de algo.

[11] Para explicar como podemos estar cientes de algo de modo inconsciente, o autor cita um artigo intitulado *Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation* [Capacidade Visual no Campo Hemianópico após uma Ablação Occipital Restrita], de 1974 (cujos autores eram neuropsicólogos membros do departamento de psicologia experimental da universidade de Oxford). Esse artigo analisa se pacientes que têm uma parte da visão comprometida por danos corticais e ainda assim percebem estímulos visuais dos quais não poderiam ter ciência, como diferenciar linhas, distinguir as letras X e O e perceber as cores vermelho e verde (Weiskrantz *et al*.1974, p. 726-727). O fenômeno é denominado de *blind-sight* [visão cega] e ilustra ao que Searle se refere quando menciona o inconsciente. Um outro exemplo de fenômeno inconsciente que ocorre conosco é o crescimento dos cabelos e unhas.

movessem tão rapidamente que a pressão interna dos movimentos da molécula nivelou-se à pressão externa do ar, cuja pressão, por sua vez, é explicada pelo movimento das moléculas pelas quais o ar exterior é composto." (Searle, 2006, p. 130). Essas duas formas de explicar atentam-se a diferentes aspectos. O primeiro modo de explicar mencionado parte do macrofenômeno para outro macrofenômeno enquanto o segundo modo parte do macrofenômeno para o microfenômeno que o causa. Teorias como doenças por micróbios ou a transmissão genética por DNA são exemplos de ciências aceitas pelo modelo de uma-coisa-menor-explica-uma-coisa-maior. Segundo Searle, isso acontece também com as células nervosas (coisa-menor) e a consciência (coisa-maior). Isso é afirmado no seguinte trecho: "grandes conjuntos de células nervosas, isto é, cérebros, causam e sustentam estados e processos conscientes" (2006, p. 132-133).

> Consciência, em resumo, é uma característica biológica de cérebros de seres humanos e determinados animais. É causada por processos neurobiológicos, e é tanto uma parte da ordem biológica natural quanto quaisquer outras características biológicas, como a fotossíntese, a digestão ou a mitose. [...] Assim que você percebe que as teorias atômica e evolutiva são fundamentais para a visão de mundo científica contemporânea, a consciência faz sentido naturalmente, como uma particularidade fenotípica evoluída de determinados tipos de organismos com sistemas nervosos altamente desenvolvidos. (Searle, 2006, p. 133-134).

Com isso, o autor argumenta que a consciência pode ser explicada causalmente por fenômenos menores que ela. Isso faz com que ele alegue ainda que seria possível recriar a consciência em um laboratório, caso soubéssemos como se dão os fundamentos que a possibilitam — com o adendo de que, para reproduzir a consciência, a artificial teria de ser parecida com as capacidades inerentes ao sistema que é capaz de causá-la[12] (ou seja, as capacidades dos neurônios).

Assim, Searle não nega a possibilidade de que uma consciência artificial possa ser criada. No entanto, ele argumenta que essa consciência não pode surgir apenas da computação, da manipulação de símbolos formais, como ocorre nos programas de computador. Para Searle, a emergência da consciência, embora ainda não totalmente explicada pela neurobiologia, depende de um sistema complexo e com uma causação essencialmente biológica. Ou seja, não basta simular processos mentais através de algoritmos e representações simbólicas; é necessário

[12] "[...] qualquer outro sistema capaz de causar consciência, porém usando mecanismos completamente diferentes, teria que ter ao menos o potencial equivalente ao do cérebro para fazer isto. (Compare: aviões não têm que ter penas para voar, mas tem realmente que compartilhar com os pássaros a capacidade causal de vencer a força da gravidade na atmosfera terrestre.)" (Searle, 2006, p. 137).

que haja uma base material adequada para que a consciência realmente se manifeste. Dessa forma, ele rejeita a ideia de que simplesmente rodar um programa em um computador possa dar origem a estados mentais genuínos, pois a consciência não é meramente uma questão de processamento de informações, mas sim uma propriedade que emerge da organização específica de sistemas biológicos. Conforme a seguir:

> [...] um modelo puramente formal nunca será, por si só, suficiente para produzir intencionalidade, pois as propriedades formais não são constitutivas da intencionalidade e não têm poderes causais, com exceção do poder de produzir o estágio seguinte do formalismo quando a máquina está rodando (Searle, 1996, p. 85).

Enfatizo, assim como o autor, o pensamento que faz jus ao termo "naturalismo biológico" para descrever sua tese, cujo objetivo é demonstrar que a consciência é uma característica biológica evolutiva que se desenvolveu nos seres capazes de sustentá-la. É uma característica biológica assim como "fotossíntese, mitose, digestão e reprodução" (Searle, 2006, p.137). Desta forma, podemos compreender o sentido do termo consciência para Searle.

Além do supracitado, o capítulo sexto de *The Rediscovery of the Mind* [A redescoberta da mente], o autor enumera doze categorias para referir-se à estrutura das propriedades da consciência, sobre as quais desejo discorrer brevemente a seguir:

> Essas categorias são coerentes entre si: estruturalidade, percepção como, a forma aspectual de toda intencionalidade, categorias e o aspecto de familiaridade. Experiências conscientes apresentam-se a nós como estruturadas, essas estruturas permitem-nos perceber coisas sob aspectos, mas esses aspectos estão sujeitos ao domínio, por nossa parte, de um conjunto de categorias, e essas categorias, sendo familiares, permitem-nos, em graus variados, assimilar nossas experiências, por mais originais que sejam, ao familiar. (Searle, 2006, p.196).

Resumidamente, desejo destacar as características estruturais da consciência, que descrevem a complexidade da experiência consciente. A primeira dessas características são as **modalidades sensoriais**, que incluem a visão, tato, olfato, paladar, audição, propriocepção e até mesmo o fluxo de pensamento. Essas modalidades compõem a base da percepção consciente do mundo e do próprio corpo.

Em seguida, destaca-se a **unidade** da consciência, que pode ser entendida sob duas perspectivas: a unidade horizontal, que armazena experiências de um curto período, como a linearidade de uma frase que eu enuncio; e a unidade vertical, que permite a percepção

simultânea de vários elementos, como a cadeira em que estou sentada, a luminária que ilumina minha escrivaninha e todas as outras coisas que me cercam percebidas simultaneamente.

A **intencionalidade** é outra propriedade fundamental, por indicar que a experiência consciente sempre possui um caráter de perspectiva. Cada pessoa tem um ponto de vista único sobre um determinado fenômeno. Esse conceito é abordado de maneira mais aprofundada por Searle em sua obra intitulada *Intentionality – An essay in the philosophy of mind*, [Intencionalidade – Um ensaio sobre filosofia da mente]. A consciência também apresenta uma **sensação subjetiva**, ou seja, cada ser sente-se de um modo particular diante de seus estados mentais. Um exemplo disso é a pergunta sobre como seria ser um golfinho, por implicar a impossibilidade de acessar diretamente a experiência subjetiva desse animal.

A consciência está intimamente **conectada** à intencionalidade de modo que, não se pode considerar apenas a intencionalidade sem atestar também a consciência. A intencionalidade verdadeira (a capacidade de a mente se referir a algo) depende da consciência e, portanto, se um sistema não possui consciência, tampouco será capaz de demonstrar intencionalidade.

A **estrutura figura-fundo** da experiência consciente é um princípio organizacional essencial. Nossa percepção se estrutura de tal forma que certos objetos estão no centro da atenção, enquanto outros formam um pano de fundo, não sendo o foco principal, mas ainda presentes na percepção. Ou seja, quando percebemos uma garrafa à nossa frente, a percebemos sempre em contraste com um fundo, e isso organiza o modo como concebemos as coisas espacialmente.

Outro aspecto importante é a **familiaridade**, que permite reconhecer categorias de objetos com os quais interagimos frequentemente. Essa capacidade de reconhecimento é fundamental para a navegação e compreensão do ambiente ao redor. O transbordamento é uma característica que indica que os estados conscientes frequentemente se referem a algo além do seu conteúdo imediato, demonstrando uma conexão com informações e experiências anteriores.

A distinção entre **centro e periferia** na consciência refere-se às características que estão no foco da atenção e aquelas que permanecem na periferia da percepção, como a sensação da cadeira tocando as costas. As condições de limite também são essenciais para a consciência, por incluírem a localização espaço-temporal e sócio-biológica dos estados conscientes. Saber onde estamos, em que época vivemos e quem somos são aspectos fundamentais para a construção da identidade e da percepção da realidade. Exemplo disso é que escrevo esse texto dando foco em minha interpretação baseada em leitura, mas perifericamente sei que o ano é 2025 e que estou no Brasil.

O **humor** é outro elemento crucial, representando o estado de espírito no qual nos encontramos, mesmo que seja neutro e não nomeável. Diferentemente da emoção, o humor é uma disposição geral que não tem intencionalidade essencial e que influencia como percebemos e interagimos com o mundo. Por fim, a consciência também possui uma dimensão de **prazer e desprazer**. Assim como os estados de humor, existem também estados de apreciação ou rejeição de experiências, que influenciam diretamente nossas ações e escolhas. Essas características estruturais são fundamentais para compreender a complexidade da experiência consciente e sua influência na percepção e interação com o mundo.

1.2.6 A subjetividade dos fenômenos mentais

Subjetividade é um termo que também pode ser empregado de várias maneiras para dizer coisas distintas. Pode ser concebida como o contrário de objetividade, o que significaria estar sujeita a relatividade dos diferentes pontos de vista. No entanto, o sentido em que a teoria searleana emprega o termo diz respeito ao modo ontológico de subjetividade[13]. Isso porque, os fenômenos que experienciamos em primeira pessoa não podem ser experienciados por terceiros de modo idêntico. Ou seja, se sentimos uma dor ou um prazer, do modo como sentimos, somente nós mesmos podemos conceber. Eis então a subjetividade. Ainda que o mundo seja objetivo, "[...] meu acesso ao mundo através de meus estados conscientes se dá sempre em perspectiva, sempre a partir de meu ponto de vista. [...] A **ontologia** do mental é uma ontologia irredutivelmente de **primeira pessoa**" (Searle, 2006, p. 140-141, grifo nosso). Adiciona:

> Se tentamos desenhar nossa própria consciência, acabamos desenhando o que quer que seja de que estejamos conscientes. [...] não podemos atingir a realidade da consciência da forma que, utilizando a consciência, podemos atingir a realidade de outros fenômenos. [...] Se tento observar a consciência de outro, o que observo não é sua subjetividade, mas simplesmente seu comportamento consciente, sua estrutura, e as relações causais entre estrutura e comportamento (Searle, 2006, p.143).

[13] Uma adição explicativa relevante para compreensão do sentido de subjetividade aqui empregado é o significado de *qualia*. "'*Qualia'* (singular: '*quale'*) é a palavra latina ('qual tipo') introduzida na filosofia contemporânea como termo de arte para se designar as qualidades fenomenais, subjetivas e conscientes da nossa vida mental, que seriam acessíveis mediante introspecção [...]. Em cada experiência que o sujeito realiza, há algo característico para ele que é realizar tal experiência. Em cada estado mental consciente, há algo característico para o sujeito que é como encontrar-se em tal estado. Há algo característico que é como sentir dor; há algo característico que é como perceber vermelho; há algo característico que é como estar apaixonado; há algo característico que é como estar deprimido etc." (Pereira, 2013, p.1)

O autor se opõe ao método fenomenológico de introspecção para investigar nossa consciência, pois, em princípio, já é uma empreitada fadada, tendo em vista que o modelo de observação que estamos acostumados a fazer não é capaz de contemplar a consciência em toda sua complexidade. “Nosso modelo moderno de realidade e da relação entre realidade e observação não pode acomodar o fenômeno da subjetividade. (Searle, 2006, p. 145).

A tentativa de representar o que é a subjetividade a limita, pois ao passo que buscamos compreender objetiva e epistemologicamente a subjetividade, a representamos conforme nossas impressões ontologicamente subjetivas. (Searle, 2006, p. 146). O modo como acreditamos que um objeto deve ser estudado nos sabota desde o primeiro momento em que tentamos, pois não há para a subjetividade a verificação nos termos ‘observador e observado’.

2 O DEBATE ATUAL SOBRE REDES NEURAIS E CONSCIÊNCIA ARTIFICIAL

2.1 Apresentação do capítulo

Sabendo que a IA tem sido um dos temas centrais do debate filosófico e científico contemporâneo, especialmente no que diz respeito à possibilidade de consciência artificial, por conta dos avanços das redes neurais artificiais, muitas teorias clássicas da filosofia da mente foram desafiadas, levando a um intenso debate sobre se essas arquiteturas podem realmente replicar ou simular a cognição humana.

Neste contexto, um dos principais pontos de discussão é a relação entre a abordagem conexionista da IA e a teoria da mente proposta por John Searle. No segundo capítulo desta monografia, explorei como o conexionismo tem sido apresentado como um modelo alternativo para entender a cognição e se ele, de fato, desafia as críticas de Searle. Seu argumento central, expresso no experimento mental do quarto chinês, sustenta que manipular símbolos formais não equivale a compreender seu significado, o que levanta dúvidas sobre a possibilidade de uma consciência artificial genuína.

Outro ponto essencial abordado é o conceito de *Background* como uma capacidade fundamental da consciência humana. Segundo Searle, a consciência depende de uma estrutura contextual que permite a inteligibilidade e a intencionalidade genuína. No entanto, as redes neurais operam a partir de padrões estatísticos, sem um contexto experiencial similar ao humano. Assim, a ausência desse *Background* pode ser apontada como um dos fatores que limitam a possibilidade de consciência artificial plena.

Dessa forma, este capítulo visou a investigar a relação entre redes neurais e consciência artificial a partir dessas perspectivas, avaliou se os avanços atuais são suficientes para superar as críticas levantadas e qual é o impacto dessas considerações para o futuro da IA. O debate sobre o potencial da consciência artificial continua em aberto, e compreender os argumentos em jogo é essencial para um posicionamento mais inteirado sobre essa questão.

2.2 O conexionismo desbanca Searle?

Conforme mencionado na introdução deste trabalho, o contexto em que Searle pensava a IA era totalmente distinto do cenário que vivenciamos atualmente. Isso pode fazer você pensar: ‘Passados 45 anos desde que Searle publicou seu artigo com o argumento do quarto

chinês, não haveria hoje condições suficientes para desbancá-lo e defender que a IA não apenas manipula símbolos sem ter compreensão semântica deles?'. Estou empenhada em verificar isso.

O argumento do quarto chinês critica os modelos computacionais baseados na máquina de Turing, cujo funcionamento era mais simples e limitado do que os modelos denominados conexionistas, por exemplo, já existentes na época de Searle, porém em nível mais hipotético por não haver ainda computadores poderosos o suficiente para servir-lhes de sede. Esses modelos trazem a promessa de fidelidade ao funcionamento da biologia humana, pois operam com um algoritmo mais robusto que funciona com base no sistema de aprendizagem profunda[14], cujo objetivo é (re)produzir o resultado das interações entre neurônios, enquanto a máquina de Turing operava apenas por meio de funções em linguagem simbólica limitada.

Veja a seguir a proposta do movimento conexionista, que:

> [...] entusiasma muitos pesquisadores por causa da analogia entre redes neurais e o cérebro. Os nós se assemelham a neurônios, enquanto as conexões entre nós se assemelham a sinapses. A modelagem conexionista, portanto, parece mais "biologicamente plausível" do que a modelagem clássica. Um modelo conexionista de um fenômeno psicológico aparentemente captura (de forma idealizada) como neurônios interconectados podem gerar o fenômeno (Rescorla, 2024, tradução nossa).

Acontece que essa empreitada tomou proporções notáveis visto que hoje interagimos com diversas de suas aplicações. Isso acarretou até aqui em:

> [...] avanços transformadores na inteligência artificial (IA), incluindo sistemas capazes de reconhecer objetos complexos em fotografias naturais tão bem ou até melhor que os humanos; derrotar grandes mestres humanos em jogos estratégicos como xadrez [...]; criar imagens e textos inéditos que, por vezes, são indistinguíveis daqueles produzidos por humanos; analisar os mais tênues ecos de rádio para descobrir novos exoplanetas orbitando estrelas a milhares de anos-luz de distância; processar enormes volumes de dados gerados por aceleradores de partículas para tentar encontrar contradições ao modelo padrão da física; [...] (Buckner, 2024, p. 2, tradução nossa).

.

O desejo de mencionar os aparatos hoje existentes se dá pelo fato de que, mesmo que utilizem outras formas de processamento de informação além da manipulação simbólica, ainda não parecem suficientemente convincentes no que diz respeito a exprimir consciência, compreensão-além-programação. Portanto, é pertinente analisar se, diante de novas

---

[14] Esse assunto está em voga e constante desenvolvimento: em 2024, o psicólogo e cientista da computação Geoffrey Hinton e o físico e biólogo John J. Hopfield ganharam o prêmio Nobel de física por contribuições para o desenvolvimento das redes neurais artificiais e *machine learning* [aprendizado de máquina]. A biografia dos premiados aponta que seus esforços iniciaram na década de 80, quando ainda havia pouco poder de computação.

possibilidades tecnológicas e aplicações possíveis, a IA ainda é limitada apenas para o que foi programada.

Essas redes neurais nas quais são baseados os programas mais recentes têm evoluído e motivam o empenho não só da área científica como também comercial. À medida que as IAs baseadas nesse modelo se mostram capazes de realizar tarefas que se assemelhem a tarefas até então delegáveis apenas a humanos, mais interessantes e populares se tornam. Podemos visualizar sua aplicação tendo diversas utilidades, como na medicina, em detecção de doenças por análise de imagens; na linguagem natural, que simula a conversa com seres humanos para deixar mais fluidos os assistentes digitais; na robótica, com máquinas autônomas que podem dirigir ou até mesmo auxiliar pessoas com condições motoras incapacitantes[15] etc.

Voltemos, então, para o sentido de compreensão no escopo searleano: é uma particularidade da cognição humana, e sua caracterização se dá da seguinte forma: "implica não só a posse de estados mentais (intencionais) como também **condições de verdade desses estados (validade, sucesso)**." (Searle, 2006, p. 70, grifo nosso). Ou seja, da mesma forma que o sujeito do quarto chinês manipula símbolos sem entender chinês; um algoritmo pode identificar um gato em uma imagem sem ter qualquer noção do que significa gato para que possa realmente validar a afirmação de que é um gato. Isso nos mostra que a evolução do programa para execução de tarefas até então inéditas não irá implicar necessariamente a emergência de consciência e de intencionalidade.

Ao falar de conexionismo é primordial mencionar os filósofos Paul e Patricia Churchland, que defendiam a ideia de que as redes neurais possibilitariam a emergência de consciência pelo seu empenho em reproduzir a cognição humana e, consequentemente, quando o avanço tanto da neurociência quanto da tecnologia permitisse, poderíamos atestar consciência a um sistema artificial. Inclusive, o casal debruça-se sobre o argumento do quarto chinês em seu artigo *Could a Machine Think?* [Uma Máquina Poderia Pensar?] de 1990, que critica fortemente a linha de raciocínio utilizada por Searle e alega que seu argumento comete petição de princípio por considerar como axioma que a sintaxe não é constitutiva nem suficiente para a semântica.

A aposta dos Churchlands é que as redes neurais podem ser a resposta para concebermos uma consciência artificial. Isso porque, enquanto a máquina de Turing "[...] é **guiada por um**

---

[15] Um campo interessante que está em desenvolvimento consiste na aplicação de interfaces cérebro-máquina (ICM), cujo objetivo é ler a "atividade elétrica [cerebral] [...], buscando, por meio de modelos computacionais, decodificá-la e interpretá-la, permitindo que se torne uma informação e gere alguma aplicação, como, por exemplo, o movimento de um membro do corpo. É como converter a intenção que está no cérebro em uma ação do mundo real [...] (Casatti, 2024).

conjunto de regras recursivamente aplicáveis que são sensíveis à identidade, à ordem e à disposição dos símbolos elementares que encontra como *input*" (Churchland; Churchland, 2015, p.158, grifo nosso), o sistema baseado por redes neurais a supera porque "[...] transforma qualquer padrão de *input* em um padrão correspondente de *output*, tal como **ditado pelo arranjo e força das muitas conexões entre os neurônios**" (Churchland; Churchland, 2015, p. 164, grifo nosso).

Essa superioridade das redes neurais, segundo os Churchlands, reside justamente em sua capacidade de processar informações de forma paralela e distribuída, diferentemente do processamento sequencial e simbólico das máquinas de Turing. Ao simular o funcionamento do cérebro humano, as redes neurais artificiais operam por meio de conexões ajustáveis que permitem aprendizado, adaptação e generalização a partir da experiência, características essenciais da cognição e, potencialmente, da consciência. Assim, os Churchlands veem nas redes neurais um caminho promissor para a construção de sistemas que não apenas executem tarefas, mas que possam vir a exibir traços de consciência.

Apesar de hoje enxergarmos seus palpites com certa plausibilidade, é importante notar que, à época em que os autores se dedicavam a essas questões, ainda não existia sequer uma máquina de Turing capaz de passar no Teste de Turing. A função que permitiria gerar um *output* suficientemente satisfatório para enganar um interlocutor humano ainda não havia sido descoberta (Churchland; Churchland, 2015, p. 158), tampouco estavam disponíveis os recursos computacionais e algorítmicos necessários para alcançar esse nível de sofisticação.

O que ansiavam era um melhor desempenho na pesquisa da IA clássica, que apontasse "[...] uma variedade de lições aprendidas com o cérebro biológico e [...] uma nova classe de modelos computacionais inspirados na sua estrutura" (Churchland; Churchland, 2015, p. 163). Nos dias atuais, isso tem sido ostensivamente investigado e as tentativas de aprimoramento e cópia das características da cognição por meio de redes neurais artificiais têm avançado significativamente.

Inspiradas na organização e no funcionamento do sistema nervoso, essas redes têm se mostrado especialmente eficazes em tarefas como reconhecimento de padrões, processamento de linguagem natural e tomada de decisão em ambientes complexos. Ainda que a analogia com o cérebro humano seja, por vezes, mais metafórica do que estrutural, a IA contemporânea tem conseguido simular aspectos do comportamento inteligente de forma cada vez mais convincente.

Entretanto, não convincentes o suficiente para Searle, uma vez que, em seu artigo, já havia resposta às objeções e, dentre elas, uma semelhante à posição dos Churchlands. Veja a seguir a objeção que visa a desbancar o argumento do quarto chinês:

> A máquina [...] simula a estrutura formal dos cérebros [...] ao processar [...] histórias e fornece respostas [...] como *outputs*. Podemos até imaginar que a máquina não opera com um único programa serial, mas com um conjunto de programas operando em paralelo, da mesma maneira que cérebros humanos possivelmente operam quando processam linguagem natural. (Searle, 1996, p. 78).

E, então, resposta à objeção, estruturada por Searle:

> A ideia central da IA no sentido forte é que não precisamos saber como o cérebro funciona para saber como a mente funciona. A hipótese básica é que existe um nível de operações mentais que consiste em processos computacionais sobre elementos formais que constitui a essência do mental e pode ser realizado através de diferentes processos cerebrais, da mesma maneira que um programa computacional pode ser rodado em diferentes hardwares. A pressuposição da IA no sentido forte é que a mente está para o cérebro assim como o programa está para o hardware, e podemos entender a mente sem fazer neurofisiologia. [...] mesmo que cheguemos a um conhecimento muito grande das operações do cérebro, isto não seria suficiente para produzir a compreensão. [...] O problema com o simulador cerebral é que ele está simulando coisas erradas acerca do cérebro. Na medida em que ele simula unicamente a estrutura formal das sequências de atividades neuronais nas sinapses, ele não está simulando o aspecto mais importante do cérebro, ou seja, suas propriedades causais e sua habilidade para produzir estados intencionais. (Searle, 1996, p. 79-80)

Em outras palavras, Searle discute a questão partindo da crítica à ideia de que um programa de computador, com os *inputs* e *outputs* certos, poderia ser consciente ou ter intencionalidade. Quando ele afirma que "O que importa nas operações do cérebro não é a sombra do formalismo dado pela sequência das sinapses, mas as propriedades efetivas de tais sequências" (Searle, 1996, p. 85), está destacando que não basta simular o funcionamento do cérebro por meio de regras formais — é preciso levar em conta as propriedades reais que existem no organismo humano e que produzem a mente.

Com base nessa ideia, na primeira parte desta monografia, busquei identificar e organizar algumas dessas propriedades reais e específicas do nosso organismo apontadas por Searle. Entender o que é exclusivo da experiência humana é essencial para que possamos definir limites claros e refletir com mais responsabilidade sobre o uso e o desenvolvimento das novas tecnologias.

A antiga aspiração de replicar — ou, ao menos, emular — certas faculdades cognitivas humanas ganhou novos contornos. A discussão que antes era predominantemente teórica se aproxima agora de desafios concretos, éticos e práticos, à medida que sistemas artificiais começam a desafiar contextos sensíveis, como medicina, justiça e relações interpessoais.

Ou seja, as intuições desses autores anteciparam debates que hoje são centrais na filosofia da mente e na ética da IA. A distância entre os modelos teóricos e a realização prática era imensa, mas isso não impediu que se formulassem questões fundamentais sobre consciência, intencionalidade e agência artificial. Não obstante, pelo impacto prático que hoje presenciamos, essas teses podem ser investigadas na busca de base para um parecer sobre o que a realidade atual nos aponta.

O risco de assumir uma posição como a dos Churchlands hoje em dia pode acarretar decisões nocivas para a própria humanidade, à medida que a confiabilidade cega — baseada na afirmação de que a IA tem uma consciência como a nossa por replicar perfeitamente os aspectos do nosso cérebro — pode implicar o sucateamento e na desvalorização da cognição humana. Não devemos relegar nossa capacidade de interpretar, julgar e ponderar às máquinas que operam com base em padrões previamente programados. Isso significaria delegar irresponsavelmente um potencial intelectual que nos é próprio e insubstituível, pelo que até aqui foi analisado.

Ter como requisito apenas o desempenho impressionante do programa não é suficiente para aceitar a IA como dotada de consciência. A verdadeira reflexão crítica e a compreensão contextual exigem mais do que processamento de dados; demandam discernimento, autonomia e interpretação de valores, elementos que as máquinas, por mais avançadas que sejam, ainda não possuem.

Em suma, o que distingue a mente humana é o fato de produzir "crenças com direcionalidade, conteúdo proposicional, condições de satisfação[16]; crenças que têm a possibilidade de ser fortes ou fracas, ansiosas ou seguras, dogmáticas, racionais ou supersticiosas, fé cega ou especulações hesitantes." (Searle, 1996, p. 76). Voltarmo-nos para essas propriedades e capacidades particulares a nós é primordial.

[16] Esse termo emprega mais uma especificidade da mente humana que a máquina não pode realizar. As condições de satisfação dizem respeito aos critérios que estabelecemos para validar um estado mental ou um ato de fala. Ou seja, são os meios pelos quais concebemos que o que está nos sendo apresentado corresponde à realidade. Ex.: "Está chovendo" só é satisfeito se realmente estiver chovendo; essa relação com o ambiente externo é a característica da intencionalidade, que é sempre sobre algo para o qual a consciência se direciona.

2.3 O *Background* como capacidade da consciência

Outro elemento conceituado por Searle que nos aponta mais uma particularidade da intencionalidade é o *Background*, termo que utilizarei, como o autor, para referir-me ao conjunto de capacidades não-representacionais que possibilitam a sustentação de condições de satisfação dos estados intencionais. Um exemplo para concebermos essa capacidade consiste em verificarmos que quando decido sentar-me em uma cadeira, essa ação não exige de mim um raciocínio consciente sobre como dobrar os joelhos ou equilibrar o corpo, pois possuo um *Background* que permite a realização dessa atividade. Essas habilidades fazem parte do *Background*, de modo que minhas ações fluam naturalmente. O autor assevera essa tese no seguinte excerto:

> É em geral impossível para os estados intencionais determinar condições de satisfação isoladamente. Para ter uma crença ou desejo, tenho que ter uma completa Rede de outras crenças e desejos. Assim, por exemplo, se agora quero comer uma boa refeição num restaurante da região, tenho que ter um grande número de outras crenças e desejos, como as crenças de que há restaurantes na vizinhança, que restaurantes são o tipo de estabelecimento onde se servem refeições, que refeições são o tipo de coisas que podem ser compradas e comidas dentro de restaurantes em determinadas horas do dia por determinadas quantias de dinheiro, e assim por diante, mais ou menos indefinidamente. (Searle, 2006, p. 250)

O *Background* é importante pois a maneira como fazemos as coisas possui mecanismos implícitos em seu acontecer, importantes para que possamos interpretar cada situação como única, de modo que ao dizer ‘traga-me um copo com água’ em um restaurante, pelo contexto, entendemos e esperamos o mesmo: que nos tragam um copo cheio de água potável e que seja entregue em mãos ou colocado sobre a mesa, e não que seja jogado em nossa cara ou que a água seja suja ou que o conteúdo despejado em nossa cabeça. O conteúdo adicional não-representacional e pré-intencional nos possibilita que tenhamos certas convicções dadas como certas. Ou seja, em todas as situações em que empregamos as expressões, temos pensamentos ou tomamos atitudes, há uma ligação entre seu significado e o *Background* que as possibilita.

> [...] [a] manifestação de minhas capacidades de *Background*, compromete-me [, por exemplo,] com a proposição de que os objetos são sólidos, ainda que eu não tenha formado nenhuma crença concernente à solidez dos objetos. [...] os sistemas de oposição quente e frio, Norte e Sul, macho e fêmea, vida e morte, Leste e Oeste, alto e baixo, etc. são todos fundamentados no *Background*. (Searle, 2006, p. 264-265)

Estamos agora diante da descrição de uma das características mais singulares do organismo humano que destoa de uma IA, pois, de acordo com essa tese, é possível a nós a compreensão das expressões não apenas sintaticamente, mas semanticamente — conforme fora dito anteriormente, impossível para a IA.

> Toda intencionalidade consciente — todo pensamento, percepção, compreensão etc. — só determina condições de satisfação relativamente a um conjunto de capacidades que não são e não poderiam ser parte desse mesmo estado consciente. Por si só, o conteúdo efetivo é insuficiente para determinar as condições de satisfação (Searle, 2006, p. 270-271).

Veja as consequências práticas dessa tese em um exemplo hipotético: em um evento, à mesa de jantar, uma pessoa diz 'Estou explodindo de tanto comer!'. Ao 'ouvir' tal afirmação, um robô prontamente acionaria um protocolo de emergência para evacuação do local, pois interpretaria a frase literalmente. Aqui, temos uma ilustração do que é não possuir a capacidade de *Background*. Em 1983, na obra *Intentionality – An essay in the philosophy of mind*, [Intencionalidade – Um ensaio sobre filosofia da mente], o autor também explicita sobre o que quer dizer ao postular o *Background*:

> [...] um conjunto de práticas, habilidades e atitudes que permitem que os estados Intencionais funcionem nas diversas maneiras que funcionam e é nesse sentido que o *Background* funciona causalmente, ao fornecer um conjunto de condições capacitantes para a operação dos estados Intencionais (Searle, 2002, p. 220).

Podemos atribuir ao *Background* nossas capacidades de compreender o significado literal mas também metafórico e possuir habilidades físicas como andar de bicicleta — em que as instruções explícitas são necessárias somente enquanto se está aprendendo e, após isso, são 'imbuídas' nos conteúdos intencionais.

CONSIDERAÇÕES FINAIS

Apesar de o cenário econômico e tecnológico que inspirou a reflexão de Searle diferir em muitos aspectos com os quais lidamos hoje, diante do avanço iminente e cada vez mais acelerado das tecnologias, dilemas morais continuam surgindo face aos novos inventos comerciais do século XXI — como assistentes virtuais que utilizam linguagem natural para elaborar respostas, criar imagens e até mesmo atuar com algoritmos para identificar perfis de potenciais consumidores. A invenção de robôs humanoides que se expressam por meio de linguagem natural, embora inicialmente cause estranheza e seja tema recorrente em produções ficcionais, torna-se cada vez mais plausível à medida que são implementados aprimoramentos tecnológicos que antes eram apenas imaginados por Searle.

A tese central do autor — de que a mente não pode ser reduzida a manipulações formais de símbolos, como defendem o funcionalismo e o computacionalismo — permanece relevante, especialmente diante dos avanços contemporâneos em IA. Ao afirmar que a consciência e a intencionalidade são propriedades emergentes de sistemas biológicos complexos, Searle estabelece uma distinção fundamental entre simulação e compreensão real, o que continua a oferecer uma base filosófica sólida para refletirmos sobre os limites das tecnologias atuais. Através do experimento mental do quarto chinês, ele demonstra que uma IA pode simular respostas inteligentes sem qualquer tipo de entendimento genuíno, o que reafirma a validade de sua crítica no contexto atual, em que máquinas se tornam cada vez mais sofisticadas.

As análises do recorte que escolhi, inseridas ao longo deste trabalho, permitiram compreender a crítica de John Searle à tese da IA Forte, com base na distinção entre sintaxe e semântica e na defesa da intencionalidade intrínseca como elemento constitutivo da mente consciente. Averiguamos que, segundo Searle, sistemas computacionais, mesmo os mais sofisticados, como redes neurais de aprendizagem profunda, não podem ser considerados genuinamente conscientes ou intencionais, pois operam unicamente sobre manipulação simbólica e não possuem acesso a um *Background* compartilhado que torne significativa sua operação.

Dessa forma, uma das principais implicações teóricas desta pesquisa é que a consciência não é uma propriedade funcional que pode ser reduzida a um sistema de regras ou simulações computacionais, mas sim uma característica emergente de certos sistemas biológicos, como o cérebro humano, em virtude de sua estrutura e causalidade interna.

Conforme a interpretação do autor que escolhi para elaborar essa monografia, toda a computação é relativa ao observador: apenas um ser dotado de intencionalidade (nós, humanos)

pode atribuir significado às operações de um sistema. Um computador não 'sabe' o que está fazendo; ele apenas executa instruções em função de algoritmos. Somos nós que interpretamos esse funcionamento.

Esse ponto levanta desdobramentos importantes para a filosofia da mente e para a ética da tecnologia. Por exemplo, a atribuição de autonomia a sistemas de IA pode ser considerada uma extrapolação perigosa, pois confere a eles um *status* ontológico que não possuem. O risco não está apenas em esperar que uma IA tome decisões corretas, mas em delegar funções humanas fundamentais — como julgar, interpretar, ponderar — a entidades que não têm acesso ao mundo da vida, ao *Background* compartilhado e à experiência subjetiva.

Do ponto de vista científico e filosófico, os resultados indicam que há limites fundamentais à simulação da mente humana. Ainda que redes neurais atuais possam superar humanos em tarefas específicas (como reconhecimento de imagens ou predição estatística), elas não demonstram compreensão real, tampouco intencionalidade genuína. Nossas habilidades oriundas do *Background* cultural, corporal e histórico devem ser conhecidas e valorizadas por nós.

Ainda, como o tempo esteve ao meu favor, pude verificar que ao longo desses anos, os argumentos de Searle encontraram respaldo em diversas evidências contemporâneas que demonstram os limites das inteligências artificiais, especialmente dos *Large Language Models* [modelos de linguagem de grande escala] (LLMs), como os utilizados por sistemas de IA generativa, que entendem e geram linguagem natural. Embora esses modelos apresentem desempenho impressionante em tarefas linguísticas, eles operam exclusivamente com base em correlações estatísticas e manipulação sintática de dados, sem qualquer tipo de compreensão semântica ou consciência do que produzem — como ilustrado no quarto chinês.

Outro fato interessante é que pesquisas com *brain organoids* [organoides cerebrais] — agregados cerebrais cultivados em laboratório a partir de células-tronco — oferecem indícios de que a consciência está profundamente ligada à estrutura biológica do cérebro[17], o que corrobora diretamente a tese de Searle. Essa descoberta vai ao encontro da ideia que o autor defende de que nossas propriedades biológicas são emergentes de nosso organismo e não podem ser explicadas apenas por manipulação simbólica ou por relações funcionais abstratas, como propõem o funcionalismo e o computacionalismo. Ele insiste que a mente só pode surgir de sistemas com as características físico-biológicas apropriadas — como o cérebro humano — e não de sistemas artificiais que apenas simulam comportamento inteligente.

---

[17] Cf. Trujillo *et al*, 2019.

Sua contribuição nos alerta sobre os perigos de confundir simulação de entendimento com compreensão real, enfatizando que a singularidade da mente humana reside em sua base biológica, a qual não pode ser replicada pelas máquinas. É primordial, então, examinar os problemas filosóficos levantados, analisando a IA a partir da hipótese segundo a qual as máquinas são limitadas em suas aspirações de alcançar compreensão genuína. Explorar parte de sua argumentação pôde fundamentar uma reflexão aprofundada sobre os avanços digitais e suas implicações para a humanidade, destacando a importância de conceber o que há de distinto do pensamento humano diante do progresso tecnológico que assimila diversas de suas características.

As discussões apresentadas ao longo desta monografia também nos direcionam para implicações práticas relevantes no campo da ética da inteligência artificial, da regulamentação tecnológica e da própria pesquisa científica. Ao distinguir claramente entre simulação computacional e experiência consciente, a tese de Searle contribui para alertar contra uma aceitação acrítica da IA como substituta da cognição humana.

No âmbito ético, isso implica reconhecer os riscos de antropomorfização de sistemas artificiais, atribuindo-lhes capacidades ou responsabilidades que não possuem. No campo jurídico e regulatório, visualizamos a necessidade de elaborar normas que não apenas garantam a transparência no uso de IAs, mas também respeitem os limites conceituais entre agentes humanos e sistemas automatizados. Por fim, na esfera científica, a valorização da base biológica da mente humana reforça a importância de investigações interdisciplinares — envolvendo neurociência, biologia e filosofia — como caminho para compreender com maior profundidade os fenômenos da consciência e da inteligência.

Saliento que ainda existem muitas questões em aberto. Searle não resolve completamente o problema de como atestar a consciência em um sistema artificial, caso isso um dia ocorra. Ele rejeita tanto o dualismo quanto o reducionismo, sustentando que a mente é um fenômeno real, emergente, mas ainda mal compreendido em termos científicos. Esse impasse aponta para a necessidade de novas investigações interdisciplinares, envolvendo neurociência, filosofia, ciência cognitiva e computação.

Em suma, os resultados desta pesquisa reforçam a posição de que, por mais avançada que seja uma IA, ela não é um sujeito consciente e não deve ser tratada como tal. Os limites apontados por Searle continuam relevantes, especialmente no cenário atual, em que sistemas como *ChatGPT*, *DALL·E* ou robôs autônomos levantam novas questões éticas e epistemológicas. Os enredos ficcionais de séries, filmes e jogos que abordam o avanço da tecnologia não me prendem a atenção apenas pelo aparente apelo à distopia, mas sim por me

parecerem plausíveis em demasia. Caberá aos futuros estudos enfrentar os desafios que encontraremos pelo caminho, com apoio da interpretação de que a mente humana não é apenas um programa executado em *hardware* biológico, mas uma realidade fenomenológica enraizada no corpo, no mundo e na cultura.

John Searle possui uma vasta composição bibliográfica e suas contribuições vão muito além das obras aqui mencionadas, que contemplam campos como a linguagem e a sociedade. Além disso, ainda que tenha me dedicado a investigar suas argumentações concernentes à filosofia da mente, o fiz a custo de um recorte significativo e, pensando à moda de Searle, estou sujeita aos aspectos da minha experiência subjetiva, que podem me privar de enriquecer com outros detalhes esse assunto que abordei. Ainda há muito para fazer nessa área que é desafiadora, que se atualiza rapidamente.

O que não se pode negligenciar é a importância da pesquisa nesse campo, que avança continuamente, impulsionada por investimentos diante do potencial lucrativo e pelo crescente interesse da população. Diante de novas gerações que crescerão tendo acesso cotidiano a ferramentas baseadas em IA, torna-se essencial que nos informemos e articulemos nosso pensamento crítico, a fim de refletir sobre os rumos que estamos tomando — e evitar que acabemos enredados na própria trama que estamos tecendo. Meu objetivo é continuar investigando quais as implicações do uso da IA e empenhar-me em levantar considerações que nos permitam ponderar sobre o grau de confiança que podemos depositar em uma tecnologia ainda experimental, profundamente marcada por interesses econômicos e com alto potencial de alienação social.

Intriga-me ainda saber as argumentações de outras obras do autor, bem como verificar no detalhe as objeções feitas e como ambos se saem quando confrontados com as tecnologias que possuímos atualmente. Gostaria de investigar como poderíamos atestar a emergência de consciência em sistemas não orgânicos uma vez que não podemos conferir outras consciências em primeira pessoa. Ou seja, se a ontologia da consciência é verificada apenas em primeira pessoa, quais os parâmetros poderíamos utilizar para verificar êxito em duplicar uma consciência artificialmente? Encerro esta monografia ciente de que não esgotei o tema, mas reconhecendo a complexidade do debate entre mente, máquina e consciência, que continua notoriamente exigindo atenção crítica e interdisciplinar.

REFERÊNCIAS

BRENTANO, F. **Psicología desde un punto de vista empírico.** Tradução de José Gaos. Madrid: Revista de Occidente, 1935, p. 85-86.

CANAL, R. Sobre as máquinas de Turing. **Perspectivas em ciências tecnológicas**. Pirassununga: Faculdade de Tecnologia, Ciência e Educação (FATECE), 2012, v. 1, n. 1, p. 22-43, fev. 2012.

CASATTI, Denise. Interface cérebro-computador é pesquisada para diagnóstico de epilepsia, depressão e outras doenças. **Jornal da USP**, 21 fev. 2024. Disponível em: https://jornal.usp.br/ciencias/interface-cerebro-computador-e-pesquisada-para-diagnostico-de-epilepsia-depressao-e-outras-doencas/. Acesso em: 28 mar. 2025.

CHURCHLAND, P. M.; CHURCHLAND, P. S. Uma máquina poderia pensar? Tradução de Nara Ebres Bachinski. **Revista Eletrônica de Filosofia**, São Paulo: Centro de Estudos de Pragmatismo, Programa de Estudos Pós-Graduados em Filosofia, Pontifícia Universidade Católica de São Paulo, v. 12, n. 1, p. 157–169, jan./jun. 2015. Disponível em: http://www.pucsp.br/pragmatismo. Acesso em: 5 abr. 2025.

DENNETT, D. C**. The intentional stance.** Cambridge: MIT Press, 1989.

FODOR, J. A. O problema mente-corpo [The mind-body problem]. **Scientific American**, v. 244, n. 1, p. 124-132, 148, 1981. Tradução de Saulo de Freitas Araujo. Reimpressão: Osvaldo Pessoa Jr. São Paulo: [s.n.], 2011. Disponível em: https://opessoa.fflch.usp.br/sites/opessoa.fflch.usp.br/files/Fodor-Port-4.pdf. Acesso em: 13 mar. 2025.

GARDENAL, I. A grelina e os paradoxos da obesidade. **Jornal da Unicamp**, ed. 184, 5 a 11 ago. São Paulo: [s.n.], 2002. p. 3. Disponível em: https://unicamp.br/unicamp/unicamp_hoje/ju/agosto2002/unihoje_ju184pag3a.html. Acesso em: 18 mar. 2025.

HUSSERL, E. **Ideias para uma fenomenologia pura e para uma filosofia fenomenológica**: introdução geral à fenomenologia pura. Prefácio de Carlos Alberto Ribeiro de Moura. Tradução de Márcio Suzuki. São Paulo: Idéias & Letras, 2006, p. 190.

NOBEL PRIZE OUTREACH. **The Nobel Prize in Physics 2024**. NobelPrize.org, 2025. Disponível em: https://www.nobelprize.org/prizes/physics/2024/summary/. Acesso em: 28 mar. 2025.

OLIVEIRA, M. A história dos primeiros anos da internet no Brasil. **Revista pesquisa FAPESP**, ed. 180, fev. São Paulo: [s.n.], 2011. Disponível em: https://revistapesquisa.fapesp.br. Acesso em: 12 mar. 2025.

PEREIRA, R. H. Qualia. **Compêndio em linha de problemas de filosofia analítica.** Lisboa: Centro de Filosofia da Universidade de Lisboa, 2013.

RESCORLA, Michael. The computational theory of mind. **The Stanford Encyclopedia of Philosophy**. Winter 2024 Edition. Disponível em:

https://plato.stanford.edu/archives/win2024/entries/computational-mind/. Acesso em: 28 mar. 2025.

ROBINSON, Howard. Dualism. **The Stanford Encyclopedia of Philosophy**, Spring 2023 Edition. Disponível em: https://plato.stanford.edu/archives/spr2023/entries/dualism/. Acesso em: 5 abr. 2025.

SEARLE, J. R. **A redescoberta da mente.** Tradução de Eduardo Pereira e Ferreira Martins. São Paulo: Martins Fontes, 2006.

SEARLE, J. R. **Intencionalidade**. Tradução de Julio Fischer, Tomás Rosa Bueno. 2. ed. São Paulo: Martins Fontes, 2002.

SEARLE, J. R. Mentes, Cérebros e Programas. Tradução de Cléa Regina de Oliveira Ribeiro. In: TEIXEIRA, J.F., **Cérebros, Máquinas e Consciência**: Uma introdução à Filosofia da Mente, São Carlos: Editora da UFSCar, 1996, p. 61-
94.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**, v. 3, n. 3, Cambridge: Cambridge University Press, 1980, p. 417-424.

SEARLE, J. R. **The rediscovery of the mind**. Cambridge: MIT Press, 1992.

TRUJILLO, C. A. *et al*. Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. **Cell Stem Cell**, v. 25, p. 558-569, 3 out. 2019. Disponível em: https://doi.org/10.1016/j.stem.2019.08.002. Acesso em: 23 mar. 2025.

TURING, A. M. Computing machinery and intelligence. **Mind**, v. 59, n. 236, p. 433-460, Oxford: Oxford University Press, 1950.

WEISKRANTZ, L. *et al*. Visual capacity in the hemianopic field following a restricted occipital ablation. **Brain**, Oxford: Oxford University Press, v. 97, 1974, p. 709–728.