

**Argumento do quarto chinês:
ausência de intencionalidade e significado na inteligência artificial**
**Chinese room argument:
absence of intentionality and meaning in artificial intelligence**

ANDRÉ RENAN BATISTELLA NOARA¹

Resumo: O presente estudo tem por objetivo desenvolver uma análise referente ao modo com o qual o filósofo americano John Rogers Searle (1932) trabalha os conceitos de intencionalidade e significado, bem como observar de que forma tais conceitos aparecem em sua crítica à Inteligência Artificial. Para tanto, será utilizado para este estudo, fundamentalmente, dois escritos de Searle, sendo estes: *Mente, Linguagem e Sociedade* (2000) e *Mentes, Cérebros e Programas* (1996) — os demais escritos de Searle serão utilizados enquanto bibliografia complementar. A metodologia abordada para o desenvolvimento do presente estudo se dará do seguinte modo: primeiramente será investigado acerca dos conceitos de intencionalidade e significado. Compreendidos tais conceitos, a investigação se voltará para compreender quais são os fundamentos que sustentam a crítica de Searle à Inteligência Artificial.

Palavras-chave: Programa. Computador. Linguagem.

Abstract: The present study aims to develop an analysis of the way in which the American philosopher John Rogers Searle (1932) works with the concepts of intentionality and meaning, as well as to observe how such concepts appear in his critique of Artificial Intelligence. For this purpose, two of Searle's works will be used for this study, namely: *Mente, Linguagem e Sociedade* (2000) and *Mentes, Cérebros e Programas* (1996) — the other works by Searle will be used as a complementary bibliography. The methodology addressed for the development of this study will be as follows: first, it will be investigated about the concepts of intentionality and meaning. After these concepts are understood, the investigation will turn to understand what are the foundations that support Searle's criticism of Artificial Intelligence.

Keywords: Program. Computer. Language.

Introdução

Mente, linguagem, intencionalidade e significado são temas que perpassam e dialogam nos escritos do filósofo americano John Rogers Searle (1932). Em sua

¹ Doutorando em Filosofia pelo Programa de Pós-Graduação em Filosofia (PPGFIL) na Universidade Federal de Santa Catarina (UFSC) na área de concentração Epistemologia e Lógica. Mestre em Filosofia pelo Programa de Pós-Graduação em Filosofia (PPGFIL) da Universidade Federal da Fronteira Sul (UFFS), campus Chapecó, na linha de pesquisa Conhecimento, Linguagem e Realidade (ano de conclusão: 2021). Licenciado em Filosofia pela Universidade Federal Da Fronteira Sul, Campus Erechim (ano de conclusão: 2018). Licenciado em Letras-Inglês pela Universidade Anhanguera (ano de conclusão: 2022). Principais áreas de interesse: Filosofia da Mente, Lógica e Filosofia da Ciência. E-mail: andre_noara@hotmail.com.

crítica à Inteligência Artificial não é diferente. Em seu artigo *Mentes, Cérebros e Programas* (1996), Searle desenvolve uma profunda argumentação em prol de demonstrar que um computador corretamente programado não pode ser considerado portador de estados mentais. Para tanto, Searle desenvolve o argumento do quarto chinês, através do qual busca demonstrar que programas de computador não possuem intencionalidade e que, por consequência, não compreendem e nem atribuem significado aos símbolos que manipulam.

O objetivo do presente estudo será analisar em que se fundamenta a crítica tecida por Searle à Inteligência Artificial. Para tanto, serão investigados dois conceitos considerados fundamentais para a discussão: intencionalidade e significado. Visando alcançar o objetivo proposto de modo preciso e elucidativo, a metodologia abordada para o desenvolvimento da presente investigação se dará da seguinte forma: primeiramente analisaremos o conceito de intencionalidade, concebido por Searle. Depois de compreendido o conceito de intencionalidade, voltaremos nossa investigação para compreendermos como a intencionalidade se relaciona com o conceito de significado. A análise desses conceitos é o que irá compor a primeira parte de nosso estudo. Havendo a compreensão dos conceitos de intencionalidade e significado, bem como da relação contida entre ambos, partiremos com nossa análise para o argumento do quarto chinês.

A investigação do argumento do quarto chinês irá compor a segunda parte de nosso estudo. Analisaremos, a partir desse experimento mental, os argumentos utilizados por John Searle em refutação à tese sustentada pela Inteligência Artificial. Veremos como os conceitos de intencionalidade e significado fazem parte da argumentação searleana.

Intencionalidade e significado

John Searle é filósofo de relevante influência nos estudos em Filosofia da Linguagem e Filosofia da Mente². O conceito de intencionalidade perpassa

² Searle dedicou boa parte de seus escritos aos temas que tratam acerca de problemas relacionados à Filosofia da Mente. Cito aqui algumas de suas principais obras a tratar desse tema: *Intencionalidade: Um Ensaio da Filosofia da Mente* (1995), publicado originalmente em (1983), *Mente, Cérebro e Ciência* (1984), *A Redescoberta da Mente* (1997), publicado originalmente em 1992, *O Mistério da Consciência* (1998), publicado originalmente em 1997, *Mente Linguagem e Sociedade* (2000), publicado originalmente

grande parte de seus escritos, tanto em Filosofia da Mente quanto em Filosofia da Linguagem³. O termo intencionalidade, em Searle, não possui o sentido que comumente empregamos em nosso cotidiano, a saber, a vinculação entre ação e intenção — ação voluntária —, mas sim ao fato de termos nossos pensamentos dirigidos a algo, àquilo que poderíamos chamar de significado ou conteúdo de uma frase ou pensamento. Para compreendermos adequadamente o conceito de significado, pensado por Searle, é de suma importância que compreendamos corretamente seu conceito de intencionalidade.

Searle divide o conceito de intencionalidade entre intencionalidade intrínseca, intencionalidade derivada e “como se”. Para explicar os conceitos, Searle utiliza os seguintes exemplos:

1. Estou com sede agora, realmente com sede, porque não tive nada para beber o dia todo.
2. Meu gramado está com sede, realmente com sede, porque não foi regado durante uma semana.
3. Em francês, “j'ai grand soif” significa “estou com muita sede”. (SEARLE, 1992, p.116-117).

A primeira sentença utilizada pelo exemplo de Searle (estou com sede agora, realmente com sede, porque não tive nada para beber o dia todo) corresponde a uma atribuição de intencionalidade intrínseca, na qual alguém, fazendo uso de um enunciado verdadeiro, atribui um estado mental real e intencional a si mesmo. Já a segunda (meu gramado está com sede, realmente com sede, porque não foi regado durante uma semana) é empregada metaforicamente. Ou seja, afirmamos que o gramado está com sede de modo meramente figurativo ou metafórico. Não há intencionalidade intrínseca em tal atribuição. Assim sendo, Searle afirma que tal atribuição é apenas “como se” houvesse intencionalidade. No terceiro caso (em francês, “j'ai grand soif” significa

em 1998, *Consciência e Linguagem* (2010), publicado originalmente em 2002 e *Mind: A Brief Introduction* (2004).

³ A Filosofia da Linguagem searlina foi bastante influenciada pelos escritos de seu mentor, John Langshaw Austin (1911-1960). Austin, em sua obra *How to Do Things With Words — Como Fazer Coisas Com Palavras —*, publicada em 1962, elaborou a teoria dos atos de fala, na qual desenvolveu, dentre outros conceitos, o conceito de atos de fala ilocucionários. John Searle concebeu e retrabalhou a teoria dos atos de fala, inicialmente proposta por Austin, em sua obra *Speech Acts: An Essay in the Philosophy of Language* (1969). A concepção de Searle acerca do conceito de intencionalidade foi, muito provavelmente, influenciada pelos escritos de Austin, em especial pelo conceito de atos de fala ilocucionários, visto que são amplamente pensados por Searle em sua teoria da realidade social e, em sua grande parte, são defendidos por Searle enquanto atos intencionais.

“estou com muita sede”) há intencionalidade, porém, a mesma não é intrínseca ao sistema. No exemplo da frase em francês, o falante de francês pode empregar tal frase para exprimir sua intencionalidade. O significado linguístico da frase corresponde a uma forma real de intencionalidade, porém não intrínseca à própria frase, mas sim derivada da intencionalidade intrínseca do falante.

A primeira e a terceira forma de intencionalidade (intrínseca e derivada) são fundamentais para compreendermos a noção de significado pensada por Searle. Desse modo, analisaremos cada uma delas de modo mais aprofundado.

Intencionalidade intrínseca

A intencionalidade intrínseca é, para Searle, uma marca dos fenômenos mentais. De acordo com Searle, ela é

[...] simplesmente o aspecto dos estados mentais pelo qual eles são praticamente dirigidos para objetos e situações outros que não eles próprios. Assim como uma flecha pode ser lançada contra um alvo e errar, ou ser lançada mesmo que não haja alvo, do mesmo modo um estado intencional pode ser dirigido a um objeto e errar de direção, ou falhar completamente porque não há objeto nenhum ali. Uma criança pode acreditar que aquele homem é o Papai Noel quando, na verdade, ele é um empregado de uma loja de departamentos, e pode-se acreditar que há fantasmas nesta casa mesmo que fantasmas não existam. (SEARLE, 2000, p. 95).

137

O conceito de intencionalidade, proposto por Searle, fica mais claro quando pensado juntamente com o conceito que ele denomina como condições de satisfação. Os estados intencionais, como crenças e desejos, possuem certas condições de satisfação. Por exemplo, podemos esperar que chova, ter medo de que chova ou acreditar que chova. Cada um desses estados possui o mesmo conteúdo (o fato de que vai chover), mas esse conteúdo é apresentado de modos intencionais diferentes. Isto é, cada um possui condições de satisfação diferentes. De acordo com Searle, possuir condições de satisfação é um aspecto geral de um grande número de estados intencionais com conteúdo proposicional. Segundo ele, [...] a chave para compreender a intencionalidade são as condições de satisfação. Um estado intencional é satisfeito se o mundo é da maneira como é representado pelo estado intencional. Crenças podem ser verdadeiras ou falsas,

desejos satisfeitos ou frustrados, intenções executadas ou não. Em cada um desses casos, o estado intencional é satisfeito ou não dependendo se há realmente uma correspondência entre o conteúdo proposicional e a realidade representada. (SEARLE, 2000, p. 131).

Percebemos, em resumo, que a intencionalidade intrínseca é o aspecto dos estados mentais pelo qual os mesmos são dirigidos a algo para além de si mesmos; dirigidos no sentido de encontrarem correspondência de suas representações para com os objetos ou situações representadas, bem como observarem se suas condições de satisfações são satisfeitas.

Intencionalidade derivada

A intencionalidade derivada pode, de certo modo, ser entendida enquanto produto da intencionalidade intrínseca. Isto é, o falante, de acordo com Searle (2000, p.131), transfere a intencionalidade intrínseca de seus pensamentos para palavras, frases, marcas, símbolos, entre outros. Quando pronunciadas de formas significativas, as frases, marcas, símbolos, etc., passam a ter uma intencionalidade derivada dos pensamentos do falante. Logo, é possível concluirmos que o conceito de significado, para Searle, nada mais é do que o conceito de intencionalidade derivada. Ou seja, nenhuma palavra, som, entre outras formas de linguagem, possui uma intencionalidade intrínseca em seu próprio sistema linguístico, mas sim uma intencionalidade derivada; derivada da intencionalidade intrínseca do falante.

Porém há mais um elemento importante presente no conceito de significado. Uma frase, para ser significativa, deve possuir aquilo que Searle descreve como intenção de significado. O conceito de intenção de significado é razoavelmente simples. Vejamos a seguinte passagem:

Suponhamos que eu esteja aprendendo alemão. Suponhamos que eu pratique a pronúncia da língua com frequência, no chuveiro, ou durante caminhadas, em dias chuvosos, dizendo “Es regnet, es regnet, es regnet”. Nesse caso, estou apenas praticando pronúncia; não quero dizer que está chovendo. Então, qual a diferença entre dizer alguma coisa e querer dizer aquilo, por um lado, e dizer alguma coisa e não querer dizer aquilo, por outro? [...] As condições de satisfação da minha intenção quando digo alguma coisa sem querer dizer aquilo são simplesmente que a

minha intenção cause um proferimento de determinado tipo, que esteja de acordo com as regras de pronúncia da língua alemã. Mas quais são as condições de satisfação da minha intenção quando realmente quero dizer o que digo? Suponhamos que tivesse aprendido um pouco de alemão e alguém me fizesse uma pergunta que sei que quer dizer “Como está o tempo hoje?” — “Wie ist das Wetter heute?” Eu responderia “Es regnet”. Agora tenho a mesma intenção que tinha antes, a de produzir um proferimento de uma frase em alemão, mas também tenho uma intenção de significado. (SEARLE, 2000, p. 132 grifo do autor).

Ou seja, a intenção de significado ocorre quando há condições de satisfação que estão para além das condições de satisfação de fazer a pronúncia de uma determinada frase. Ou seja, a intenção de significado é equivalente à intenção de que as condições de satisfação, isto é, o próprio proferimento, tenham condições de satisfação. O conceito de condições de satisfação, juntamente com o conceito de intencionalidade derivada, são a chave para compreendermos o conceito de significado em Searle.

Compreendido em que consistem os conceitos de intencionalidade e significado, vejamos agora um argumento de grande influência dentro dos campos de estudos da Filosofia da Linguagem e da Filosofia da Mente, desenvolvido por Searle com objetivo de mostrar que não há intencionalidade ou atribuição de significado na Inteligência artificial.

139

O argumento do quarto chinês

O argumento do quarto chinês é possivelmente o mais famoso e influente dos argumentos já desenvolvidos pelo filósofo americano John Rogers Searle. Searle, em seu artigo intitulado *Mentes, Cérebros e Programas* (1996), originalmente publicado em 1981, desenvolve um experimento mental pelo qual visa lançar uma crítica aos defensores da Inteligência Artificial (IA). Porém vale ressaltar que a crítica searleana não se estende a toda a IA, mas apenas aos que defendem uma IA no sentido forte.

Searle divide a IA em forte e fraca. A IA forte argumenta em prol da ideia de que computadores adequadamente programados possuem estados cognitivos, e que os programas que as máquinas instanciam são teorias psicológicas. Isto é, para os defensores da IA forte, um programa em atividade deve ser considerado

uma mente. Já a IA fraca não é tão radical, visto que argumenta que os programas de computadores apenas podem simular estados cognitivos, mas não se constituem enquanto estados cognitivos⁴.

A crítica searlina se contrapõe radicalmente à tese defendida pela IA forte. Para Searle, um programa de computador trabalha unicamente com a manipulação de símbolos formais, os quais, para o computador programado, não possuem significado algum. Desse modo, Searle argumenta que um programa de computador corretamente programado não pode se constituir enquanto uma mente. Para melhor argumentar em prol de sua crítica, Searle desenvolve o argumento do quarto chinês, o qual será apresentado aqui de modo integral, visto que, apesar de se constituir enquanto uma passagem longa, é de suma importância para o nosso estudo:

Suponha que estou trancado em um quarto e suponha que me dão um calhamaço de papel com um texto em chinês. Além disso, suponha que eu não conheça o idioma chinês, nem escrito nem falado, e que eu não seja sequer capaz de reconhecer a escrita chinesa, ou seja, distingui-la, por exemplo, da escrita japonesa ou de rabiscos sem significado. Suponha, agora, que além deste primeiro calhamaço fornecem-me - também em chinês - um segundo, contendo um roteiro com um conjunto de regras para correlacionar o segundo texto com o primeiro. As regras são em inglês e eu as compreendo tão bem como qualquer outro falante nativo de inglês. Isto me possibilita relacionar um conjunto de símbolos formais com o outro, e o que entendo por formal aqui é que posso identificar os símbolos por seu formato. Nestas circunstâncias, imagine também que me forneçam um terceiro calhamaço contendo símbolos em chinês junto com algumas instruções, outra vez em inglês, as quais me possibilitaram correlacionar elementos deste terceiro maço com os dois primeiros; estas regras me instruem a como relacionar determinados símbolos em chinês com certos tipos de configuração e os devolver como resposta a determinadas configurações dadas no terceiro calhamaço. Sem que eu saiba, as pessoas que me fornecem os textos com os referidos símbolos, denominam o primeiro bloco de "roteiro", o segundo, de "história" e o terceiro de "questões". Ademais, eles intitulam os símbolos devolvidos em resposta ao terceiro maço de "respostas às questões", e o conjunto de regras em inglês de "programa".

⁴ Essa distinção operada por Searle pode ser melhor observada no capítulo IX da obra *A Redescoberta da Mente* (1997), intitulado *A Crítica da Razão Cognitiva*, no qual John Searle desenvolve uma investigação aprofundada em torno da ciência cognitiva e da inteligência artificial.

Para complicar a história um pouquinho mais, imagine que estas pessoas também me forneçam histórias em inglês, as quais eu compreendo, e então elas me fazem questões em inglês sobre estas histórias, e eu as devolvo respondendo em inglês. Suponha, ainda, que depois de um tempo eu me saia tão bem ao seguir as instruções para manipulação dos símbolos em chinês e que os programadores consigam escrever tão bem os programas que do ponto de vista externo - isto é, do ponto de vista de alguém que esteja do lado de fora do quarto no qual eu estou trancado - minhas respostas às questões são indistinguíveis de falantes nativos de chinês. Ninguém observando minhas respostas pode dizer que eu não falo uma palavra de chinês. Vamos também supor que minhas respostas às questões em inglês são indistinguíveis de outro falante nativo de inglês - pela simples razão de que eu sou um falante nativo de inglês. Do ponto de vista externo, - na visão de alguém que lê minhas respostas, - as respostas em chinês e em inglês são igualmente satisfatórias. Mas no caso do idioma chinês, eu obtenho respostas manipulando símbolos formais em chinês, sem significação. No que diz respeito ao chinês, eu simplesmente me comportei como um computador; executei operações computacionais com base em elementos formalmente especificados. Para os propósitos do idioma chinês, eu sou simplesmente uma instânciação de um programa de computador. (SEARLE, 1996, p. 66).

Ou seja, Searle visa demonstrar, a partir desse experimento mental, que um programa de computador apenas manipula símbolos formais, os quais, para o programa, não possuem significado algum. Searle demonstra as diferenças contidas entre a mera manipulação de símbolos e a atribuição de significado. No caso do exemplo, aquele que está trancafiado no quarto possui, quando manipulando símbolos em chinês, inputs e outputs do mesmo modo que um programa de computador possui, e sua significação dos inputs e outputs é completamente nula, do mesmo modo que para um programa de computador também o é. Porém quando esse mesmo alguém está recebendo e enviando informações em uma língua conhecida, ocorre algo que no caso dos inputs e outputs em chinês não ocorre, algo que está para além da mera manipulação de símbolos: há significado.

O argumento do quarto chinês, em resumo, é desenvolvido por Searle com vista a criticar a IA forte, ao demonstrar que programas de computador funcionam estritamente com a manipulação de símbolos formais; que tais programas não possuem intencionalidade e nem atribuem significado aos

símbolos que manipulam. O ponto central do argumento de Searle é demonstrar que a simples manipulação de símbolos formais não garante intencionalidade; um programa de computador é algo puramente formal, algo que os estados intencionais não o são. Como observamos na seção anterior, os estados intencionais não são definidos pela sua forma, mas pelo seu conteúdo; como um determinado conteúdo mental com condições de satisfação.

A intencionalidade é um dos elementos que diferencia os fenômenos mentais dos demais fenômenos e propriedades naturais do mundo. Um termostato, um vulcão, um girassol, um trovão, nenhum deles possui o que Searle entende por intencionalidade intrínseca, visto que nenhum deles possui estados mentais. Isto é, há, para Searle, uma conexão entre o mental e a intencionalidade, de tal modo que se, e somente se, houver o mental, haverá intencionalidade⁵. Searle defende que certos organismos possuem estados intencionais intrínsecos, sendo estes causados por processos próprios ao sistema nervoso desses organismos — o que não é o caso de um programa de computador.

Tal qual o girassol, o termostato e o trovão, os programas de computadores também não possuem intencionalidade. Por não possuírem intencionalidade, os programas estão restritos a funcionar de modo puramente sintático, algo que a mente transcende: com a intencionalidade, a mente humana é capaz de compreender o significado das sentenças e, por consequência, reconhecer sua semântica. Por estar restrito à sintaxe, um programa de computador não possui a capacidade de atribuir significado. Atribuir significado é pressuposto para a compreensão. Logo, a compreensão de um programa de computador é completamente nula. Um programa de computador não possui a capacidade de compreender, visto que desconhece por completo o significado das sentenças.

Poderíamos concluir afirmando que um programa de computador possui, quando muito, a forma de intencionalidade que Searle denomina por como se. Ou seja, uma intencionalidade que, em certos casos, atribuímos de modo

⁵ “[...] somente um ser que pudesse ter estados intencionais conscientes poderia ter estados intencionais de algum modo, e todo estado intencional inconsciente é pelo menos potencialmente consciente. Esta tese tem consequências enormes para o estudo da mente. Implica, por exemplo, que qualquer discussão de intencionalidade que deixe de lado a questão da consciência será incompleta”. (SEARLE, 1992, p. 190).

meramente metafórico ou figurativo, tal qual quando afirmamos que o gramado está com sede ou que a porta do shopping sabe quando fechar. Analogamente, poderíamos dizer, metaforicamente, que um programa de computador compreende tanto o significado de suas sentenças quanto um termostato compreende a temperatura do ambiente.

Considerações finais

Faremos aqui uma breve retomada dos principais pontos que investigamos no presente estudo, o qual dividiu-se em duas partes: na primeira parte, investigamos acerca dos conceitos de intencionalidade e significado; na segunda parte, concluímos a ausência de tais conceitos na inteligência artificial.

Percebemos, durante a primeira parte, que Searle divide o conceito de intencionalidade em intencionalidade intrínseca, intencionalidade derivada e “como se”. Percebemos que o conceito de significado, em Searle, refere-se ao conceito de intencionalidade derivada. Observamos, ainda, que para uma frase ser significativa, ela precisa possuir intenção de significado — o falante precisa ter a intenção de impor condições de satisfação a condições de satisfação. Entendido tais conceitos, avançamos para a segunda parte do estudo.

Na segunda parte de nosso estudo, cientes dos conceitos de significado e intencionalidade, passamos a analisar o argumento do quarto chinês. Em relação a esse argumento, percebemos, primeiramente, que o mesmo foi elaborado por Searle em crítica à Inteligência Artificial no sentido forte — a qual afirma que um computador corretamente programado constitui uma mente. A partir da análise desse experimento mental, percebemos, por analogia, que um programa de computador funciona unicamente a partir da manipulação de símbolos formais, algo estritamente sintático e desprovido de intencionalidade. Por não possuírem intencionalidade, percebemos que os programas de computadores não atribuem significado aos símbolos que manipulam, visto que a significação é derivada da intencionalidade intrínseca. Por não atribuírem significação aos símbolos, concluímos, em concordância com Searle, que um programa de computador não possui compreensão alguma; não há nível algum de significação ou compreensão do programa de computador para com os símbolos que o mesmo manipula.

Em conclusão ao presente estudo, é possível afirmarmos que o argumento do quarto chinês é válido e suficiente para demonstrar que a Inteligência Artificial forte está equivocada. A inexistência de intencionalidade e a não atribuição de significado demonstram a ausência de compreensão e estados mentais em programas computacionais.

Referências

- AUSTIN, J. L. *How to do things with words*. New York: New York Press, 1965.
- SEARLE, R.J. *Consciência e linguagem*. São Paulo: Martins Fontes, 2010.
- SEARLE, R.J. *Mind: a brief introduction*. Oxford: Oxford University Press, 2004.
- SEARLE, R.J. *Mente, linguagem e sociedade: filosofia no mundo real*. Rio de Janeiro: Rocco, 2000.
- SEARLE, R.J. *O mistério da consciência*. São Paulo: Paz e Terra, 1998.
- SEARLE, R.J. *A redescoberta da mente*. São Paulo: Martins Fontes, 1997.
- SEARLE, R.J. *Mentes, cérebros e programas*. Trad. Cléa Regina de Oliveira, In: TEIXEIRA, J. F. *Cérebros, máquinas e consciência: uma introdução à filosofia da mente*. São Carlos: SP: EDUFSCar, 1996.
- SEARLE, R.J. *Intencionalidade*. São Paulo: Martins Fontes, 1995.
- SEARLE, R.J. *Mente, cérebro e ciência*. Lisboa: Edições 70, 1984.
- SEARLE, R.J. *Speech acts: An essay in the philosophy of language*. Cambridge, Cambridge University Press, 1969.

Submissão: 10. 06. 2023 / Aceite: 30. 08. 2023