

Postado em 11 de Maio de 2021 - 16:46

## Aspectos filosóficos da Inteligência Artificial

Por Gisele Leite.

Fonte: [Gisele Leite](#)

A proposta de investigar o conceito de inteligência artificial através da filosofia contemporânea nos conduz a diversas respostas a respeito da relação corpo e mente.

O uso de computadores e outros artefatos tecnológicos na educação não se resume à mera substituição do livro impresso pelo livro digital ou ao uso de páginas de internet e correio eletrônico em cursos à distância.

A questão sobre a informática aplicada à educação visa propiciar um ambiente de ensino virtual que possa ter um alto grau de interação com os estudantes.

Enfim, para galgar tal objetivo é necessário o uso de inteligência artificial, em especial, quanto à criação de tutores inteligentes[1] e ainda a produção de softwares educacionais cada vez mais interativos.

Primeiramente, o conceito de inteligência artificial é bastante evidente e preciso nas ciências da computação, contudo, possui implicações que vão além de seu uso técnico, pois aponta para a possibilidade de se criar pensamento nas máquinas, isto é, a mente artificial.

Eis que precisamos também definir o que seja pensamento. O filósofo John Searle[2] que se dedicou ao tema, demonstra o quão controversa é a questão, e ao citar o teórico que cunhou a expressão "inteligência artificial" John McCarthy:

“O meu termostato tem três crenças - está demasiado quente aqui, está demasiado frio aqui e está bem aqui - se, ter crenças é pensar; então, o termostato pensa?”

Searle realiza distinção entre dois projetos diferentes de inteligência artificial: o projeto fraco, onde o computador é visto apenas como uma ferramenta que pode imitar alguns comportamentos racionais; e o forte, cujo objetivo é produzir uma mente artificial a partir do computador.

Podemos enquadrar o uso da inteligência artificial na confecção de programas de computador mais interativos e de tutores inteligentes na primeira categoria, o que diminuiria em muito as limitações que podemos ter quanto ao sentido da expressão inteligência artificial. Pois este se transformaria em metáfora.

O problema é que ao usar a metáfora de maneira intuitiva, podemos tomá-la no sentido que esta não tem: de ser a representação fiel de algo real.

Além do mais, a inteligência artificial aplicada à educação tem nos seus horizontes o objetivo de transformar o programa de computador em um professor virtual que fará mais que apenas imitar, com muitas limitações, o professor humano. Mas é inevitável perguntar, qual é afinal a diferença entre a mente virtual e a mente real?

Qual é a distância existente entre o imitar o pensamento e realmente pensar?

Qual é a diferença entre um computador que imita o enxadrista humano e o jogador humano?

Não seria verdade que também o enxadrista humano consegue aprender a jogar através da imitação de outros jogadores?

A imitação seria mesmo um veículo honesto de aprendizagem? E se um computador fizesse o mesmo, por que não podemos cogitar que ele está aprendendo e falando? E, essa fala implica em pensar, então por que não acreditar que a máquina está pensando?

Todas essas questões foram formuladas e respondidas pelo idealizador da inteligência artificial, Allan Turing[3], em seu texto intitulado “*Computing machinery and intelligence*” que foi publicado na revista de filosofia (Mind), o que por si só, aponta que a filosofia integra o caminho argumentativo percorrido por essa temática.

Analisando a proposta de Turing e a réplica de Searle e, ainda, a tréplica de Daniel Dennet. O que nos permite a visualizar uma parte central do debate filosófico sobre a inteligência artificial.

Outro aspecto menos considerado na questão é a relação existente entre mente e matéria. O projeto de construção de mente artificial pressupõe a relação intrínseca entre a mente e a matéria, em outras palavras, o cérebro produz a mente. Assim, o cérebro artificial produzirá a mente artificial.

A relação entre mente e corpo é complexa, posto que existem bons argumentos contra este pressuposto, ou seja, contra o materialismo reducionista, a teoria segundo a qual a mente pode ser inteiramente reduzida ao corpo.

Logo o primeiro argumento dirigido contra essa doutrina advém da concepção dualista, que defende a tese da existência de abismo intransponível entre mente e matéria.

Não se trata do dualismo clássico e cartesiano, onde a mente é considerada uma substância, mas o dualismo contemporâneo, que não nega a produção da mente pelo cérebro, mas afirma que disto não se segue que os eventos mentais possam reduzidos a meros eventos físicos.

A réplica epistemológica do materialismo reducionista e a tréplica monológica feita pelos defensores do materialismo não-reducionista. Evidentemente não se pretende apresentar resposta definitiva à questão, mas apenas dimensionar sua complexidade para propiciar uma discussão teórica em torno do tema.

A primeira questão a responder é: os computadores podem pensar? O mais adequado meio de respondê-la seria construir um computador que realizasse tal tarefa. O que seria uma típica solução empírica para resolver o problema. O idealizador de computadores e primeiro teórico a sugerir a possibilidade de se criar uma inteligência artificial, Allan Turing (1990) propôs o seguinte teste:

Algumas pessoas, os juízes, farão perguntas, via terminal do computador, a uma entidade que eles não sabem se é um computador ou uma pessoa.

Numa outra sala estarão as pessoas conectadas a uma parte dos terminais do computador com os quais os juízes irão dialogar, a outra a parte dos terminais será conectada a computadores.

Caberá aos juízes determinar se estão dialogando com uma pessoa ou com uma máquina, a tarefa dos computadores é exatamente confundir os juízes para que pensem que estão falando com uma pessoa, ao invés de uma máquina.

Segundo Allan Turing, se um computador conseguir imitar o comportamento verbal humano ao ponto de confundir os juízes, então a máquina estaria pensando.

A questão é determinar ao certo se imitar de modo perfeito o pensamento não é propriamente pensar.

Ou seja, qual é a diferença entre pensar como um ser humano e, simplesmente, pensar? Turing supõe não haver diferença nenhuma e, assim, acredita que o computador que passe pelo seu teste, que seja confundido com um ser humano, realmente é uma máquina pensante.

O referido teste elaborado por Turing causa duas impressões imediatas, a saber: a primeira é grande simplicidade, mas parece ser evidentemente implausível.

De que se possa, apenas baseado num jogo de perguntas e respostas, distinguir as respostas dadas por uma máquina, das respostas oferecidas por um ser humano. Não se segue que a máquina pense, no mínimo, seriam necessárias algumas premissas extras para se chegar à esta conclusão.

Allan Turing, evidentemente, antecipou as dificuldades de aceitação de seu teste, pois já vigiam objeções à noção de que a máquina poderia pensar. É relevante não se deter nessas críticas e respostas.

A primeira crítica apresentada refere-se às emoções, é apenas quando uma máquina tiver os mesmos sentimentos que os seres humanos têm ao produzirem um poema, por exemplo, que ela realmente pensará.

A resposta de Turing não é direta, e vem primeiro nos lembrar que usamos a prova oral com a finalidade de saber se o estudante realmente escrevera o trabalho apresentado para a banca examinadora (como no caso de mestrado e doutorado) e, fazendo isso, através de perguntas que explorem o sentido e o conteúdo semântico das palavras que estão envolvidas neste assunto.

Turing afirma que, caso um computador se comporte do mesmo jeito, não há de se supor que não esteja se comportando do mesmo modo que os seres humanos se comportam quando realizam essa prova.

Existe outra objeção que ele nomeia como Argumento a partir da informalidade do comportamento. A objeção consiste basicamente na constatação de que não é possível produzir um conjunto de regras que possa descrever o que um homem deveria fazer em cada circunstância concebível.

Disto se conclui que nós não podemos ser máquinas. A resposta de Turing é verdadeira [verdadeira] mas indica que no caso de computadores não é possível prever em detalhes o seu comportamento.

Outro artigo de Turing publicado em 1950, numa época que não havia máquinas capazes de participar deste teste. É claro que, com o desenvolvimento dos computadores surgiram máquinas e programas que podem competir no jogo da imitação.

O filósofo Paul Churchland (1996) descreve a sua participação, como juiz, no teste de Turing, que anualmente, é organizado pelo *Cambridge Center for Behavioral Studies* de Massachusetts. o Prêmio Loeber, que é conferido ao programa de computador que melhor imitar um ser humano.

E, nesse concurso, os computadores e pessoas precisam responder sobre apenas um tópico, por exemplo, futebol, e os interrogadores dão notas a cada candidato, sem saber se o candidato é homem ou máquina, no quesito de semelhança com o comportamento humano.

Os interrogadores ficam numa sala, diante de vários terminais de computadores e cada um fará perguntas para todos os terminais, com o tempo de quinze minutos por terminal.

Numa outra sala estão terminais onde há computadores e seres humanos respondendo às perguntas. Em cada sala, existem juízes que irão atestar se o processo está sendo feito dentro das regras. Churchland foi um destes juízes e diz que computadores já conseguiram enganar os interrogadores, mas que, quando foi juiz, nenhum dos interrogadores foi enganado pelas máquinas, entretanto, cinco dos oito interrogadores, tomaram um ser humano por computador!

Turing, em seu artigo, afirma ser muito fácil um ser humano imitar um computador, mas Churchland diz que este é um dos problemas do teste: seres humanos não são muito hábeis, dentro das condições do teste, em distinguir outros seres humanos de máquinas.

### **A sala chinesa**

As críticas que o teste de Turing recebeu não referem, obviamente, a sua exequibilidade empírica - experimento é perfeitamente realizável, mas ao que é o pressuposto pelo argumento: que basta uma máquina ter um comportamento visível idêntico ao de um ser humano quando este exerce uma atividade que requer pensamento (falar) para que constata-se que a máquina pensa.

O filósofo John Searle (1990) propôs um experimento mental para refutar[4] o tipo de experimento empírico proposto por Turing.

O experimento ficou famoso e chamado de argumento da sala chinesa e, talvez seja o experimento mental mais difundido em toda a discussão a respeito da possível inteligência artificial.

Nesse experimento, o filósofo Searle, que desconhece completamente o idioma chinês, está trancado numa sala onde recebe uma folha com um grupo de caracteres em chinês, depois lhe é fornecido uma

segunda folha também com caracteres em chinês, acompanhada de outro papel com regras em inglês, língua que ele entende e domina, para relacionar os símbolos da segunda folha com os da primeira folha.

As regras escritas em inglês lhe informam que toda vez que na primeira linha da primeira folha ocorre determinado grupo de símbolos e na primeira linha da segunda folha, outro determinado grupo de símbolos, ele deverá escrever uma terceira folha, um grupo de símbolos, tudo em chinês e deverá passar esta terceira folha por uma janela para alguém do lado de fora, observe que ele irá identificar os símbolos exclusivamente pelos seus desenhos, ou seja, pela sua forma.

No experimento mental, Searle recebe as folhas em chinês por uma janela e entrega a folha que escreveu, por outra janela.

Para quem está do lado de fora da sala, as folhas que são inseridas nesta [sala] contém um texto em chinês e perguntas sobre este texto e, a folha que Searle tem dentro da sala, retorna ao exterior que contém as respostas às perguntas, de tal forma que quem souber chinês entenderá perfeitamente estas respostas.

Para quem sabe chinês, a sala estará falando chinês, já que pode responder perguntas que lhe são feitas, porém, e Searle, sabe chinês? É claro que não, pois tudo o que está fazendo é manipular os símbolos cujo significado é desconhecido.

Searle utilizou o experimento mental para demonstrar que, mesmo no caso do experimento empírico proposto por Allan Turing dar o resultado esperado, não provará que os computadores estão realmente se comunicando e, portanto, que estão de fato pensando, porque tudo o que o computador fará é seguir as regras formais do seu programa, onde símbolos são substituídos por outros símbolos, mas disto não se segue que compreenda o significado dos símbolos.

Segundo Searle, *in litteris*:

"Essencial à nossa concepção de um computador digital é que as suas operações possam ser especificadas em termos puramente formais ou sintáticos, é fatal para a concepção de que os processos mentais e os processos de programa são idênticos. (...) A razão por que nenhum programa de computador pode alguma vez ser uma mente é simplesmente porque um programa de computador é apenas sintático, e as mentes humanas são mais do que sintáticas. As mentes são semânticas, no sentido de que possuem mais do que uma estrutura formal, têm um conteúdo".

O principal argumento do filósofo tem como pressuposto primacial a concepção da mente humana e, a fortiori, do pensamento, como aquilo que, além de ter habilidades[5] sintáticas, necessariamente possui habilidades semânticas.

Outro pressuposto é o de que as habilidades semânticas não podem ser derivadas, ou, pelo menos, exclusivamente derivadas, de habilidades sintáticas. Neste pressuposto está envolvido o conceito de intencionalidade: nossas habilidades semânticas[6] envolvem necessariamente a intencionalidade e (...) nenhum modelo puramente formal jamais será suficiente por si mesmo para a intencionalidade porque as propriedades formais não são por si mesmas constitutivas da intencionalidade (...).

O terceiro pressuposto é o de que os computadores realizam atividades que são puramente formais, ou, conforme ele descreve, podem ser especificadas em termos puramente formais.

O argumento de sala chinesa recebeu e ainda recebe inúmeras críticas, e o próprio Searle imitando Turing, compilou e responde a estas críticas. Mas enfocando apenas o segundo pressuposto, aquele que envolve o conceito de intencionalidade.

É relevante sublinhar, no entanto, que, até onde sabemos, nenhuma crítica foi feita ao primeiro pressuposto, aquele que afirma que o pensamento envolve necessariamente a capacidade semântica.

É verdade que muitas críticas, contudo, foram feitas ao terceiro pressuposto, visto que muitos teóricos julgaram que os programas de computadores não realizam atividades puramente formais.

A intencionalidade é conceito medieval que fora redescoberto pela filosofia e psicologia, particularmente pelo alemão Franz Brentano[7], do século XIX. O termo advém do verbo latino *intendo* que significa apontar, indicar.

Brentano utilizou o conceito para definir algo que é típica e exclusivamente mental: apenas os fenômenos mentais indicam alguma coisa, representam algo, o que não ocorre com os fenômenos físicos. Imagine um pedaço de chumbo, ele não representa nada em si mesmo.

Imagine, agora, que uma pessoa veja este pedaço de chumbo e perceba que há uma palavra escrita nele; a mente desta pessoa irá conectar os símbolos (as letras) que estão impressas no chumbo com aquilo que significam, ou seja, com aquilo que apontam ou indicam; a palavra é lida por alguém que possa compreendê-la que a intencionalidade ocorre, o pedaço de chumbo com a palavra, se não é lido por ninguém, não passa de um pedaço mudo de matéria.

Na medida em que a intencionalidade é a capacidade que a mente tem de representar objetos, nossa habilidade semântica é intencional.

Porém, a intencionalidade não se refere apenas a compreender o significado de uma palavra a partir do significado de outras palavras, conforme ocorre no dicionário, pois implica ter crenças e desejos com relação à palavra. E ter uma crença ou desejo não é algo puramente formal.

Ter a crença de que está chovendo, é enfim, ter um conteúdo mental, e ter certos desejos relativos a ela, o desejo de se molhar, ou desejo de pegar um guarda-chuva.

Portanto, não basta dotar o computador de um dicionário ou de uma enciclopédia para que possua intencionalidade, uma vez que tudo o que fará é substituir um símbolo (por exemplo, as letras da palavra, chuva) por outros símbolos como as letras da expressão “precipitação atmosférica formada de gotas d'água”, mas não existirão crenças nem desejos relacionados a estes símbolos.

Searle (1997) apresenta um outro argumento que poderíamos chamar de o uso ambíguo do conceito de intencionalidade.

Relembremos a surpresa de Searle ao constatar que, para alguns teóricos, os termostatos possuem crenças, com intuito de entender este tipo de afirmação, ele criou a distinção entre a intencionalidade original e a intencionalidade derivada.

A intencionalidade original pode ser ilustrada pela frase: Estou com sede, já a intencionalidade derivada é exemplificada por outra frase como: O gramado está com sede.

É claro, que nesse segundo caso, cogita-se metaforicamente, ou como o filósofo prefere, está implícito o como se: faz tempo que o gramado não é regado, as gramas estão amareladas, logo pode-se afirmar que é como se o gramado tivesse o desejo por água. Apenas seres humanos e animais podem desejar e, é somente com relação a estes que se pode se ter a intencionalidade original.

O filósofo defende o uso ambíguo do termo intencionalidade quando este se refere às máquinas, neste caso, trata-se apenas da intencionalidade derivada.

O programa de computador possuiria uma intencionalidade derivada da intencionalidade original do ser humano que o criou. Mas há uma tese implícita nesta observação: a intencionalidade derivada jamais se transformará numa intencionalidade original.

É uma tese óbvia se pensarmos que a intencionalidade derivada nada mais é do que uma metáfora. Assim, para que máquinas possam ser dotadas de intencionalidade é preciso eliminar a diferença entre a intencionalidade original e a intencionalidade derivada, essa é a estratégia adotada por Daniel Dennett[8], *in verbis*:

"A solução para o problema da nossa intencionalidade é direta. Nós simplesmente concordamos que os artefatos representacionais (como descrições escritas e esboços) possuem intencionalidade derivada em virtude do papel que desempenham nas atividades de seus criadores.

Uma lista de compras escrita em um pedaço de papel possui apenas a intencionalidade derivada que obtém do agente que a escreveu.

Bem, da mesma forma a lista de compras mantida pelo mesmo agente na memória"

Ela é interna (neste último caso), não externa, mas ainda assim, é um artefato criado pelo seu cérebro e significa o que significa em razão da posição particular na economia em funcionamento das atividades internas de seu cérebro e do papel que exerce no controle das atividades complexas do seu corpo no mundo real que nos cerca."

A teoria de Dennett parece ser muito contra-intuitiva, pois nega a diferença entre ter um conteúdo na mente e ter o mesmo conteúdo fora da mente. E, para o filósofo, a intencionalidade da mente humana é derivada da evolução que criou nossos corpos, sendo resultante do desenvolvimento de intencionalidades[9] primitivas.

Portanto, nossa intencionalidade é uma intencionalidade derivada, e não, original, conforme pensava Searle.

Ora, se nossa intencionalidade é derivada e, ainda assim, nós pensamos, então por que um computador, cuja intencionalidade é nitidamente derivada da nossa, não poderia também pensar?

É claro que um computador precisa fazer mais do que substituir símbolos para demonstrar intencionalidade, é necessário que demonstre capacidade de interagir com o mundo físico; por esta razão, Dennett propõe que o computador seja colocado num robô que possa se relacionar com o mundo, tanto no sentido de perceber o que o cerca (adaptando câmeras de TV e outros sensores no robô), quanto no de modificar o seu ambiente (construindo braços mecânicos para o robô e um meio de locomoção).

Além do aspecto mecânico, é fundamental que o computador-robô não seja inteiramente programado no que diz respeito às suas ações: o robô deverá realizar certas tarefas especificadas no programa, também terá regras que o orientarão na realização de suas tarefas, mas deverá aprender como o ambiente, através da tentativa e do erro a melhor forma de realizar sua tarefa.

O nosso robô terá duas características que aproximam a sua intencionalidade derivada daquela dos seres humanos:

1. deverá interagir fisicamente com o ambiente;
2. e, deverá ser capaz de aprender a partir desta interação.

É interessante notar que o escrito de Turing já cogitava de máquinas que aprendem, e ele indica que talvez este seja o melhor caminho para produzir a inteligência artificial.

A tecnologia contemporânea está bastante desenvolvida com relação à construção de robôs, e existe o objetivo de criar robôs não apenas capazes de interagir com o ambiente como também aprendam através desta interação.

Nesse sentido, parece apenas uma questão de tempo o desenvolvimento de um robô que possa demonstrar comportamento intencional e, por conseguinte, pensamento[10].

Já é o momento adequado para se avaliar o esclarecimento através de reflexões filosóficas que proporcionem uma análise crítica e lúcida quanto ao uso da inteligência artificial na educação.

A primeira tentativa de prova empírica de que uma máquina pode pensar. Está relacionada com o programa de computador que pode imitar a capacidade linguística de um ser humano. E, já existem programas que podem realizar esta tarefa, mas vimos, através do argumento da sala chinesa, que a prova empírica não é suficiente.

Surge então, um novo tipo de prova empírica: um computador com corpo, um robô, que pode aprender através da interação com o meio, com os seres humanos e, quem sabe, possa se criar uma sociedade de robôs, o que replicaria todo o ambiente no qual as mentes humanas existem.

É claro que muitas críticas foram novamente feitas a esta nova tentativa empírica de comprovar a inteligência artificial.

O próprio Searle e Dennett teceram argumentos contra esta possibilidade.

A criação de uma mente artificial implica a criação de um artefato que produza esta mente - seja computador digital, um robô ou um sistema conexionista, por trás desta ideia está o pressuposto de que a mente é o resultado de um processo físico que ocorre em nosso cérebro (ou em todo nosso corpo) e que, como qualquer processo físico, pode ser reproduzido em um sistema físico, que seja organizado de tal modo a obter este resultado.

Se a mente pode ser produzida por um sistema constituído por neuroproteínas, por que não poderia ser produzida por um sistema feito de metal e silício?

Em nossa vigente concepção científica de mundo, esta pergunta soa como puramente retórica, pois é evidente que, se a matéria, numa determinada organização, cria a mente, então basta reproduzir esta organização para se criar uma mente artificial.

O problema é que desde o início da história da filosofia existem teorias que negam a identidade entre mente e matéria.

Entre os primeiros filósofos encontra-se Anaxágoras cuja cosmologia afirmava que "todas as coisas estavam unidas e imóveis por um período infinito de tempo, e que a Mente introduziu movimento e a separou".

Para Anaxágoras, portanto, é a mente quem ordena a matéria. Já outro filósofo pré-socrático, Demócrito, identificava mente com corpo, sendo um dos fundadores do materialismo[11].

Atomista, ele acreditava que a mente era composta de átomos, na verdade, átomos esféricos, que também constituíram fogo e eram mais propícios para causar o movimento.

A filosofia moderna começa com a disputa entre um filósofo materialista, Thomas Hobbes, e o pai do dualismo moderno, Descartes.

Não apenas o materialismo chegou até nossos dias, existem versões mais contemporâneas do dualismo e concepções materialistas que negam a possibilidade de ser reduzir a mente à matéria. Iremos analisar brevemente estas concepções para saber em que sendo elas estão relacionadas com o projeto da inteligência artificial.

Mas, não apenas o materialismo chegou até nossos dias, como também existem versões contemporâneas do dualismo e concepções materialistas que negam a possibilidade de se reduzir a mente à matéria.

### **O dualismo contemporâneo**

O dualismo é hoje uma posição minoritária na discussão filosófica contemporânea sobre a relação entre a mente e o corpo.

Ainda assim, o número de teses e argumentos dualistas é muito grande para serem analisados. Ficou bastante conhecido na filosofia da mente contemporânea: “*What is it like to be a bat*”, do filósofo norte-americano Thomas Nagel[12] (1974).

O escrito principia questionando a tese materialista de que todo evento ou fenômeno mental pode ser reduzido a um evento ou fenômeno físico.

Quando a ciência explica um fenômeno, ela o transforma em um fenômeno científico, por exemplo: quando a ciência explica que a água é composta de dois átomos de hidrogênio e um átomo de oxigênio, ela está reduzindo a substância água para a substância H<sub>2</sub>O.

O mesmo ocorre ou pode ocorrer com os fenômenos mentais? Pois podemos reduzi-los a fenômenos físicos sem deixar resto?

Nagel também recorre a um experimento mental. E, pede que o leitor se coloque no lugar de um morcego. Como é perceber o mundo através do sonar natural dos morcegos? É óbvio que a experiência sensorial que os morcegos possuem não pode ser imaginada por nós.

Agora, vamos mudar o experimento. Imaginemos extraterrestres que também percebem luz visível, mas são incapazes de perceber as cores, como eles poderão imaginar a nossa sensação de cor? Note que a cor é fenômeno físico e que pode ser percebido por um ponto de vista externo: tanto nós humanos quanto os animais que possuem visão e mesmo possíveis extraterrestres têm acesso a este fenômeno, em outras palavras, ele é objetivo.

Já que a percepção [que] temos das cores, ou a percepção auditiva do morcego, pertence a um ponto de vista interno, e portanto, único, pois apenas os morcegos percebem as coisas do modo como os morcegos as percebem, apenas nós percebemos as cores da maneira como as percebemos.

O que experimento visa a estabelecer é uma irreducibilidade do subjetivo ao objetivo. Nagel não nega que a experiência subjetiva, enfim, nossa consciência, não se relacione com o mundo, o que nega é a possibilidade de explicar a nossa consciência a partir de fenômenos físicos, porque esta explicação deixaria algo de fora - a redução não seria completa.

Ele não está defendendo a tese cartesiana de que a mente é uma substância distinta do corpo, apenas sustenta que não podemos explicar a mente a partir do corpo.

Que lições podemos tirar do dualismo contemporâneo para o projeto da inteligência artificial? Conforme vimos, morcegos percebem o mundo, mas nós somos incapazes de imaginar como é perceber o mundo no modo de morcegos.

Podemos construir um robô que seja capaz de perceber o mundo, mas isto jamais nos ajudará a saber como nós percebemos o mundo, pois o robô irá perceber o mundo à maneira dele.

Ora, como um dos objetivos da inteligência artificial é explicar, através da simulação de nosso pensamento, em máquinas, a cognição humana, este objetivo parece não ser realizável, se Nagel está certo.

Por outro viés, quando o dualismo estabelece uma distância intransponível entre a mente e a matéria (o que não é exatamente o caso do dualismo de Nagel), mas certamente é o caso de outros defensores contemporâneos do dualismo.

Então, apesar de existir uma ligação entre mente e matéria, jamais seremos capazes de perceber qual é esta ligação, o que implica a incapacidade de saber que tipo de configuração da matéria causa a mente - uma pedra de cal no projeto da inteligência artificial.

Evidentemente que para os defensores do materialismo que realizaram inúmeras respostas ao desafio dualista. Cabe destacar a resposta oferecida por David Lewis porque apresenta um argumento claro e bem estruturado, além de ter influenciado profundamente a discussão sobre este tema.

Lewis, em primeiro lugar, define a teoria da identidade psicofísica: A teoria da identidade psicofísica é a hipótese de que não por necessidade, mas por uma questão de fato - toda a experiência (eventos mentais) é idêntica com algum tipo de estado físico.

Em outras palavras, a teoria da identidade psicofísica afirma que a cada evento mental corresponde a um evento físico.

Para provar tal princípio é necessário partir de premissas que não o pressupõem, e, Lewis aponta duas premissas que julga auto-evidentes.

A força da primeira reside em ser uma mera análise do conceito de experiência: A primeira das minhas duas premissas para estabelecer a teoria da identidade é o princípio de que a característica definitiva de qualquer experiência como tal é a causalidade (causal role).

Podemos chamar esta premissa de princípio empirista, sempre que há experiência, há relações causais.

Outra premissa, que julga ser uma hipótese plausível é a de (...) que existe um corpo unificado de teorias científicas que nós aceitamos e que fornece uma explicação completa e verdadeira de todo fenômeno físico (isto é, todo o fenômeno que pode ser descrito em termos físicos).

A segunda premissa é a ideia de que a ciência consegue dar uma explicação satisfatória da realidade (aqui entendida como conjunto de fenômenos físicos).

Existe, é evidente, uma observação extra: todas as nossas experiências mentais, estão associadas a experiências físicas, ou seja, não existe mente sem corpo. Ora pela primeira premissa, temos que toda experiência deve implicar em uma relação causal, assim, a relação entre a experiência mental e a experiência física deve ser uma relação causal.

Como as ciências têm a capacidade de explicar as relações causais entre os fenômenos físicos, nada impede que elas possam explicar as relações causais entre fenômenos físicos e mentais.

O filósofo admite que possam existir fenômenos que não sejam físicos, mas como não têm poder causal - pois, caso tivessem esse poder, seriam explicados pela ciência- então eles não constituem experiência, uma vez que a primeira premissa afirma que toda experiência deve ter poder causal.

O materialismo propõe um dilema ao dualismo. A tese dualista é a de que os fenômenos mentais não podem ser explicados ou reduzidos pelos fenômenos físicos. Os dualistas precisam admitir, entretanto, que os fenômenos mentais possuem poderes causais, caso contrário, a mente seria inerte.

O problema é que as relações causais são definidas pela maneira como ocorrem no mundo físico e cabe à ciência investigá-las.

É claro que esta, em seu estágio atual, não consegue explicar completamente a relação entre os fenômenos mentais e os fenômenos físicos, mas o dualista vai além: afirma que ela nunca será capaz de explicar esta relação.

Adicionando a premissa de que tudo o que pode ser experienciado é algo que admite relações causais, então sobram duas alternativas para o dualista: ou concede que os eventos mentais não sejam passíveis de relações causais, tornando-os não experienciáveis, ou admite que sejam passíveis de relações causais.

Mas, nesse caso, nada impede que possam ser causalmente relacionados a fenômenos físicos, a não ser que existam num outro tipo de realidade, uma espécie de realidade não-física.

Enfim, postular uma realidade deste tipo acarreta o ônus de provar a sua existência.

### **O Monismo anômalo**

A teoria que vamos examinar não tenta dissolver o dilema apresentado mais acima, porque não defende o dualismo.

Porém, apesar de não ser dualista, tampouco é materialista no sentido estrito, visto que nega a tese de que todo evento ou fenômeno mental pode ser reduzido a um evento ou fenômeno físico.

O monismo anômalo defende uma concepção que à primeira vista parece contraditória:

1. eventos mentais interagem casualmente com eventos físicos;
2. toda integração causal pode ser expressa por uma lei;
3. não existem leis que conectem os eventos mentais aos eventos físicos e vice-versa.

Segundo a teoria, os eventos mentais são idênticos a eventos físicos. E, por isto, ela é monista, que é o oposto do dualismo, mas disto não se segue que eles possam ser reduzidos aos eventos físicos, porque para que isto ocorra é preciso que os eventos mentais sejam subsumidos às leis que se aplicam aos eventos físicos, o que a teoria nega ser possível - daí o adjetivo anômalo, ou seja, aquilo que não segue leis.

O monismo anômalo foi proposto pelo filósofo Donald Davidson. Não há espaço para expor em detalhe seus argumentos, assim, tudo o que podemos fazer é um esboço da teoria.

Uma lei da natureza ou a lei científica que explique os eventos mentais a partir dos eventos físicos é uma lei psicofísica.

Davidson precisa provar que não existem leis psicofísicas. Leis se aplicam a objetos que possuem afinidade categorial: eu não posso explicar leis civis a objetos inanimados, do mesmo modo, não posso aplicar leis físicas aos números (não faz sentido eu dizer que o número três é mais leve que o quatro).

É evidente que podemos aplicar leis físicas ao corpo humano, afinal, ele é um objeto físico, mas podemos aplicar leis físicas à mente humana?

Davidson argumenta que:

"Do mesmo modo que a satisfação das condições de medição da massa ou da extensão podem ser vistas como constitutivas do campo de aplicação das ciências que aplicam estas mediações, a satisfação das condições de consistência e coerência racional pode ser vistas como constitutivas do campo de aplicação de conceitos como crença, desejo, intenção e ação".

Em síntese, as categorias que a física utiliza incluem os conceitos como massa, extensão, velocidade, mas não incluem conceitos referentes à racionalidade: quando explica-se por que fulano decidiu sair levando um guarda-chuva, apesar de não estar chovendo, uso categorias como a crença que ele tem na previsão do tempo apresentada na televisão e o desejo de não se molhar - são categorias que não tem correspondência no mundo físico.

É constitutivo da explicação de nossas ações que eu possa me referir ao desejo que originou a ação e a crença que guiou este desejo (no caso, a ação é pegar o guarda-chuva, o desejo é o de não se molhar e a crença é a de que hoje choverá).

Visto que as categorias que usamos para explicar os eventos mentais não são as mesmas que a física utiliza para explicar os eventos físicos, então a resposta à pergunta necessariamente é negativa: o campo categorial dos eventos físicos não é o mesmo que o dos eventos mentais; logo não se podem aplicar as leis físicas aos eventos mentais.

É relevante sublinhar, contudo, que isso não impede que os eventos mentais sejam idênticos a eventos físicos, apenas não se pode estabelecer leis causais entre estes, assim, existe um campo do mental que é irreduzível ao que é físico.

Não é tão fácil saber o que o monismo anômalo acarreta para o projeto da inteligência artificial. Em primeiro lugar, é uma teoria materialista, portanto, defende a tese de que o mental é idêntico ao físico, uma tese que implica a possibilidade de se criar um mecanismo que possa produzir a mente.

Por outro lado, ao afirmar que as leis não se aplicam ao fenômeno mental, ele impede o conhecimento de quais as leis físicas causam o surgimento da mente.

Assim, mesmo que possam construir uma máquina que pense, analisar o seu funcionamento em nada nos ajudará a compreender com a mente funciona.

## **Conclusão**

Enfim, o projeto de criar uma inteligência artificial é um empreendimento empírico, mas tem pressupostos que transcendem a investigação empírica.

O primeiro pressuposto é a própria definição do que seja pensamento. O que está em jogo ora é saber se uma definição meramente behaviorista esgota o sentido deste conceito.

É evidente que o pensamento está ligado à ação, mas daí se segue que toda ação seja ligada a um pensamento?

No caso do teste de Turing, é verdadeiro que toda vez que um ser humano fala ele está pensando, mas disto pode inferir que tudo o que fala pensa? Esta objeção também se estende ao robô, que pode demonstrar comportamento similar ao ser humano, e, ainda, assim, não pensar.

A objeção, contudo, não tem força lógica para eliminar a possibilidade da inteligência artificial. Se é verdade que uma máquina [pode] ter comportamento igual ao de um ser humano quando este age não implica que a máquina pensa, tampouco a possibilidade de que ela pense está excluída.

O que Searle e Nagel parecem estar afirmando é que esta máquina, para pensar, necessariamente, deverá ter vida interior, ou seja, possuir subjetividade.

De qualquer modo, precisaremos ir além do behaviorismo para determinar se uma máquina é capaz de pensar.

Existe, contudo, um pressuposto mais primordial: a relação entre corpo e mente é intrínseca ou extrínseca?

A tese dualista afirma que é extrínseca, ou seja, os predicados mentais são independentes dos predicados físicos.

Disto se segue que não podemos saber se uma determinada característica (predicado) de um sistema físico corresponde a uma determinada característica de um sistema mental.

O dualismo não exclui a possibilidade de um sistema físico qualquer (seja o cérebro, seja um computador ou um robô) produzir a mente, mas ela exclui a possibilidade de entendermos como o sistema físico está relacionado com a mente. Isto implica que podemos replicar o cérebro humano num outro meio físico - metais e silício - sem com isto replicarmos a mente humana.

A concepção materialista é a que mais favorece a possibilidade da criação de uma mente artificial, pois reduz o corpo à mente e, assim, se conseguirmos replicar com sucesso o cérebro humano em outro meio físico não há nada que impeça o surgimento de uma mente a partir deste meio físico.

O materialismo não-reducionista e monismo anômalo põe um freio no otimismo triunfante do materialismo tradicional

Segundo nossa interpretação, o monismo anômalo representa um obstáculo ao projeto de inteligência artificial similar ao do dualismo, pois cria uma barreira, neste caso, epistemológica - entre a mente e a matéria; as leis que guiam o comportamento da mente não são as mesmas que guiam o comportamento da matéria.

Ora, quando construímos um cérebro artificial estamos seguindo as leis físicas, caso estas sejam diferentes das leis mentais, então não poderemos saber se do funcionamento físico do cérebro resultará uma mente artificial.

No que tange à educação é relevante estar atento à distinção para não confundirmos um programa bastante interativo com um professor artificial, ou ainda, um programa que imita a atividade docente, por um verdadeiro docente.

Por outro lado, tampouco podemos negar a possibilidade de que um dia tenhamos computadores e robôs que possam substituir funções e profissões humanas, restando aos humanos apenas programar tais artefatos.

### **Referências:**

CHURCHLAND, Paul. The engine of reason, the seat of soul. Massachusetts. MIT Press, 1996

DAVIDSON, Donald. Essays on actions and events. Oxford university Press, 1969.

DENNETT, Daniel. Tipos de Mente. Rio de Janeiro: Rocco, 1987.

LEWIS, David. An Argument for the identity theory, with addenda . In: ROSENTHAL, David (ed0. Materialismo and the mind-body problem. Indianapolis: Hackett, 2000.

NAGEL, Thomas. What is it like to be bat? The Psilosophical Review, Vol.83, nº4, 1974.

PORTO, Leonardo Sartori. Uma investigação filosófica sobre a inteligência artificial. Revista Informática na Educação. Teoria e prática. volume 9, n.1. jan/jun. Porto Alegre: 2006.

SEARLE, John. Mente, cérebro, ciência. Lisboa: Edições 70, 1997.

TURING, Allan. Computing machinery and intelligence.Oxford :Oxford University Presse, 1990.

WITTGENSTEIN, Ludwig. Investigações Filosóficas. Lisboa: Fundação Calouste Gulbenkian, 1987.

### **Notas:**

[1] Pesquisadores fundamentados apontam que a aprendizagem privada com um tutor humano é quatro vezes mais eficiente do que a aprendizagem na sala de aula tradicional. E, nesse sentido o tutor eletrônico que consiga imitar o tutor humano pode até ser mais eficiente, pelo menos em alguns aspectos, do que as aulas tradicionais.

[2] John Rogers Searle (1932) é um filósofo e escritor norte-americano, professor da Universidade de Berkeley, na Califórnia nos EUA. Ele é membro da Academia Americana de Artes e Ciências e da Academia Europeia de Ciências e Arte, destinatário de oito títulos honoríficos e, é membro da Guggenheim Fellow, conferencista da BBC Reith e duas vezes por nomeado Fulbright Fellow. Searle começou sua filosofia com o estudo do campo da linguagem em Atos da fala, o passo inicial em uma longa viagem ainda inacabada, abraçando não só a língua, mas também nos domínios da consciência e dos estados mentais, da realidade social e institucional, da racionalidade, da conexão do "eu" (self) com a intencionalidade individual e coletiva, da percepção e do realismo direto e, mais

recentemente, na busca de uma explicação de uma estrutura racional como base para a existência de livre-arbítrio na filosofia da mente e na filosofia da sociedade.

[3] Allan Mathison Turing (1912-1954) foi matemático britânico, pioneiro da computação e considerado o pai da ciência computacional e da inteligência artificial. Durante a Segunda Guerra Mundial, Turing trabalhou para a inteligência britânica em Bletchley Park, num centro especializado em quebra de códigos. Por um tempo ele foi chefe do Hut 8, a seção responsável pela criptoanálise da frota naval alemã. Planejou uma série de técnicas para quebrar os códigos alemães, incluindo o método da bomba eletromecânica, uma máquina eletromecânica que poderia encontrar definições para a máquina Enigma.

[4] Entre os diversos textos que expressam tal ponto de vista temos o de Margaret Boden, de Cole e o Daniel Dennet. Aliás, pesquisadores do laboratório de Lira da Universidade de Genova criaram um robô que imita o comportamento de um bebê quando este percebe o mundo. O babybot fora programado com uma lista mínima de instruções a fim de que, a partir da interação com o meio exterior possa aprender a perceber os objetos.

[5] Simplificadamente, podemos dizer que habilidade é a capacidade que uma pessoa tem de realizar algo. É a facilidade, ou o dom, de proceder diante de uma tarefa qualquer. Assim, quanto mais fácil você consegue lidar com uma exigência, mais habilidoso você é.

Nesse sentido, um exemplo fácil de ser entendido é a sua habilidade de dirigir. Você pode, ou não, ter habilidades para ligar o carro sem deixar o motor morrer, mudar a marcha no tempo certo, trabalhar com os pedais, estacionar, e realizar outras ações.

E todas essas são técnicas individuais, que você aprende e pode ir treinando e aperfeiçoando até que se tornem praticamente automatizadas. Então, quando isso acontece, você é tido como uma pessoa hábil nesse tipo de procedimento.

Já a competência, também de acordo com o Aurélio, significa: direito, faculdade legal que um funcionário ou um tribunal têm de apreciar e julgar um pleito ou questão. Capacidade, suficiência (fundada em aptidão). Atribuições. Porfia entre os que pretendem suplantar-se mutuamente.

Para entender isso melhor, continuemos no exemplo do carro. Sabendo dirigir, ou seja, reunindo as habilidades necessárias para realizar essa ação, você já pode ir à rua, onde passará por situações que exigirão de você a capacidade para lidar com problemas diferentes.

Um outro carro pode te fechar, a chuva pode dificultar sua direção ou um acidente pode fazer com que você tenha que mudar seus planos. Enfim: para sair dessas situações, será necessário utilizar o conjunto das habilidades que você tem, certo?

[6] Segundo a Talbu, a aquisição e o desenvolvimento de repertórios verbais dependem de um conjunto amplo de habilidades cognitivas e sociocognitivas humanas, envolvendo compreensão da ação intencional e compreensão da intenção comunicativa. A primeira é biologicamente herdada dos primatas não-humanos (Tomasello; Tomasello et al.); a segunda, socialmente estabelecida pela interação de indivíduos linguística e simbolicamente competentes (Tomasello). Portanto, a linguagem verbal é um processo tanto biológico quanto social.

[7] Franz Clemens Honoratus Hermann Brentano (1838-1917) foi um filósofo alemão. Sua filosofia evoluiu em direção de um aristotelismo moderno, nitidamente empírico em seus métodos e princípios.

Os trabalhos mais importantes de Brentano são no campo da psicologia, por ele definida como ciência dos fenômenos psíquicos (ou, o que para ele é sinônimo, da consciência). Os objetos de seus estudos não foram, porém, os estados, mas sim os atos e processos psíquicos. Segundo Brentano, o fenômeno psíquico distingue-se dos demais por sua propriedade de referir-se a um objeto, bem como a um conteúdo de consciência, através de mecanismos puramente mentais. À psicologia caberia, então, estudar as diversas maneiras pelas quais a consciência institui suas relações para com os objetos existentes nela mesma, descrever a natureza desta relação, bem como o modo de existência deste objeto. Brentano distingue três classes de fenômenos psíquicos fundamentais: a representação, julgamento e sentimento de amor (aprovação) ou sentimento de ódio (desaprovação).

Seu trabalho mais importante publicado em vida foi *Psychologie von Empirischem Standpunkt* (Psicologia segundo o ponto de vista empírico), de 1874. A segunda edição desta obra inclui ainda *Von der Klassifikation der Psychischen Phänomene* (Sobre a classificação dos fenômenos psíquicos).

Sua obra póstuma mais importante é "*Vom sinnlichen und noetischen Bewusstsein*" (Sobre a consciência sensória e noética), de 1928. Esta obra foi publicada como terceira parte em algumas das edições posteriores da *Psicologia do ponto de vista empírico*. Ela apresenta o modo como o conceito brentano de intencionalidade foi reformulado, após a edição de 1874. Foi o mestre de filosofia de Husserl.

[8] Daniel Clement Dennett (1942) é um proeminente filósofo norte-americano. As pesquisas de Dennett se prendem principalmente à filosofia da mente (relacionada à ciência cognitiva) e da biologia. Dennett é ainda um dos mais proeminentes ateus da atualidade. Para Dennett, os estados interiores de consciência não existem. Em outras palavras, aquilo que ele chama de "teatro cartesiano", isto é, um local no cérebro onde se processaria a consciência, não existe, pois, admitir isto seria concordar com uma noção de intencionalidade intrínseca. Para ele a consciência não se dá em uma área específica do cérebro, como já dito, mas em uma sequência de inputs e outputs que formam uma cadeia por onde a informação se move. Um dos livros de Dennett é *A Ideia Perigosa de Darwin*. Em *Consciousness explained*, Dennett tem interesse na capacidade de evolução para explicar algumas das características de produção de conteúdos de consciência já é aparente, e isso se tornou uma parte integrante do seu programa. Ele defende uma teoria conhecida por alguns como darwinismo neural. Ele também apresenta um argumento contra o qual ia, ele argumenta que o conceito é tão confuso que não pode ser posto a qualquer uso ou entendido em qualquer forma não-contraditória e, portanto, não constitui uma válida refutação de Fisicalismo. Muito do trabalho de Dennett desde a década de 1990 tem se preocupado em dar as suas ideias anteriores, abordando os mesmos temas a partir de uma perspectiva evolucionária, o que distingue humanos a partir de cérebros de mentes animais (*Tipos de Mentes*), a forma livre e espontânea vontade é compatível com uma visão naturalista do mundo (*Liberdade evolui*). Em 2006 seu livro *Breaking the Spell*, Dennett tenta dar à crença religiosa o mesmo tratamento, explicando possíveis razões evolutivas para o fenômeno da adesão religiosa.

[9] Intencionalidade é um conceito filosófico recuperado por Franz Brentano da Escolástica, uma subcategoria dentro da filosofia medieval para definir o estatuto da consciência, qualificada por estar dirigida para algo, ou de ser acerca de algo, possuída pela maior parte dos nossos estados conscientes. O termo foi mais tarde usado por Edmund Husserl, que defendeu que a consciência é sempre intencional. A intencionalidade distingue a propriedade do fenômeno mental: ser necessariamente dirigido para um objeto, seja real ou imaginário. É neste sentido, e na fenomenologia de Husserl, que este termo é usado na filosofia contemporânea.

As nossas crenças, pensamentos, anseios, desejos, são sempre acerca de alguma coisa. Do mesmo modo, as palavras que usamos para exprimir essas crenças e outros estados mentais são acerca de coisas. O problema da intencionalidade consiste na compreensão da relação que se verifica entre um estado mental, ou a sua expressão, e as coisas acerca das quais esse estado mental se constitui como tal.

[10] Pensamento e pensar são, respectivamente, uma forma de processo mental ou faculdade do sistema mental. Pensar permite aos seres modelarem sua percepção do mundo ao redor de si, e com isso lidar com ele de uma forma efetiva e de acordo com suas metas, planos e desejos. Palavras que se referem a conceitos e processos similares incluem cognição, senciência, consciência, ideia e imaginação. O pensamento é considerado a expressão mais "palpável" do espírito humano, pois através de imagens e ideias revela justamente a vontade deste. O pensamento é fundamental no processo de aprendizagem (vide Piaget).

O pensamento é construto e construtivo do conhecimento. O principal veículo do processo de conscientização é o pensamento. A atividade de pensar confere ao homem "asas" para mover-se no mundo e "raízes" para aprofundar-se na realidade.

[11] O reconhecido como fundador da filosofia ocidental, Tales de Mileto, acreditava que o universo estava cheio de espíritos e que mesmo a matéria que julgamos inanimada possuía mente - era desta forma que ele explicava o movimento dos magnetos. Relata-se que Tales era tão observador da natureza que conseguiu formular estudos profundos no entendimento do mundo meteorológico, prevendo com exatidão temperaturas, climas e chuvas, orientando a prática do desenvolvimento da agricultura.

O grande mérito de Tales, na verdade, não foi a sua explicação aquática da realidade: foi o fato de que, pela primeira vez na história, o homem buscava uma explicação totalmente racional para o seu mundo, deixando de lado a interferência dos deuses. Tales pode ser tido também como o pai da filosofia unitarista — que busca a explicação de todas as coisas a partir de um único princípio (no caso dele, a água) — e que teria seu maior expoente na figura de Heráclito de Éfeso.

[12] Thomas Nagel nasceu em 04 de julho de 1938 em Belgrado, a antiga Iugoslávia, é atualmente professor de Filosofia e Direito na Universidade de Nova Iorque. Seus trabalhos se concentram em filosofia da mente, filosofia política e ética. É conhecido por sua crítica aos estudos reducionistas sobre a mente em seu "*What Is it Like to Be a Bat?*" ("Como é ser um morcego?"), de 1974, e por sua contribuição à teoria político-moral liberal e deontológica em "*A Possibilidade do Altruísmo*", de 1970. Também é autor de "*Uma breve introdução à Filosofia*", obra de divulgação que aborda temas da filosofia em linguagem acessível, propondo uma reflexão individual acerca de questões como saber o que é o certo e o que é o errado.

---



**Gisele Leite**

Professora Universitária. Pedagoga e advogada. Mestre em Direito. Mestre em Filosofia. Doutora em Direito. Conselheira do INPJ. Instituto Nacional de Pesquisas Jurídicas. Consultora Jurídica.

---

**Palavras-chave:** Aspectos Filosóficos Inteligência Artificial, Sala Chinesa, Dualismo Contemporâneo, Monismo Anômalo

FONTE: <https://www.jornaljurid.com.br/colunas/gisele-leite/aspectos-filosoficos-da-inteligencia-artificial>

Artigo capturado em 11/03/2023 13:00 h