



*Paakat: Revista de Tecnología y Sociedad*  
e-ISSN: 2007-3607  
Universidad de Guadalajara  
Sistema de Universidad Virtual  
México  
paakat@udgvirtual.udg.mx

Año 13, número 25, septiembre 2023-febrero 2024

## **Inteligência artificial fraca e força bruta computacional II**

### ***Weak artificial intelligence and computational brute force II***

Alexandre Quaresma\*

<https://orcid.org/0000-0002-9868-808X>

Pontifícia Universidade Católica de São Paulo, Brasil

Ítalo Santiago Vega\*\*

<https://orcid.org/0000-0001-8133-5221>

Pontifícia Universidade Católica de São Paulo, Brasil

[Recibido: 22/08/2022. Aceptado para su publicación: 24/05/2023]

DOI: <http://dx.doi.org/10.32870/Pk.a13n25.773>

#### **Resumo**

O objetivo deste ensaio é refletir criticamente sobre o estado atual dos nossos sistemas de inteligência artificial (IA) de última geração, as suas potencialidades latentes, as suas possibilidades técnicas promissoras e também as suas limitações lógicas, estruturais e funcionais intrínsecas. Para o efeito, tomaremos como exemplo emblemático e objeto de estudo nesta ocasião o novo LaMDA da Google, que é um poderoso sistema pericial de IA que processa a linguagem verbal falada, ou seja, um sistema concebido e utilizado de acordo com a teoria da IA fraca, que exclui - desde o início e por si mesma - qualquer hipótese plausível de consciência.

#### **Palavras chave**

Conceito de moralidade; máquina; ferramenta; filosofia da mente.

**Abstract**

*The objective of this essay is to critically reflect upon the current condition of our most modern artificial intelligence (AI) systems, their latent potentials, their promising technical possibilities, as well as their intrinsic logical, structural, and functional limitations. In order to accomplish this, we will take as an emblematic example and subject of study the new LaMDA by Google, which is a powerful expert artificial intelligence system that processes spoken language, thus a system designed and utilized under the weak theory of AI, which inherently excludes any plausible hypothesis of consciousness right from the outset.*

**Keywords**

*Moral concepts; machine tools; philosophy of mind.*

**Resumen**

*El propósito de este ensayo es reflexionar críticamente sobre la condición actual de nuestros sistemas de inteligencia artificial (IA) más modernos, sus potencialidades latentes, sus posibilidades técnicas prometedoras, y también sobre sus limitaciones lógicas, estructurales y funcionales intrínsecas. Para ello, tomaremos como ejemplo emblemático y objeto de estudio en esta ocasión el nuevo LaMDA de Google, que es un potente sistema experto de inteligencia artificial que procesa el lenguaje verbal hablado, por tanto, un sistema concebido y utilizado según la teoría débil de IA, que excluye —desde el principio y por sí misma— cualquier hipótesis plausible de conciencia.*

**Palabras clave**

*Concepto de moral; máquina; herramienta; filosofía del espíritu.*

**Introdução**

Antes de enfrentar o tema propriamente dito da inteligência artificial do LaMDA do Google e suas intrínsecas limitações, faz-se necessário afirmar, com toda segurança e logo de saída, que o fenômeno da mente consciente acontece única e exclusivamente na biologia, pelo menos até hoje, sem quaisquer outras exceções conhecidas. E se, e quando, chegasse o dia em que –devido a alguma engenhosidade certamente humana– a vida viesse a eclodir em meio cibernético-informacional, como pretendem alguns –ou seja, num computador, androide ou robô–, certamente um tempo ne-oparadigmático e inaugurador se descortinaria, significando consequências e desdobramentos importantes para a própria humanidade que cria e usa tais sistemas.

E tal feito seria de fato e incontestavelmente uma realização tecnocientífica extraordinária, magnífica, digna não apenas de nota, mas também de atenção e cuidado, enfim, de prudência e vigilância, bem como de estudos e pesquisas extensivas que pudessem se dedicar à sua plena compreensão e manutenção. “Se tudo der certo” –prevê Russell (2021, p. 17)– “esse feito anunciará a chegada de uma Idade de Ouro para a humanidade, mas precisamos levar em conta que estamos planejando construir entidades muito mais poderosas do que os seres humanos”.

A pergunta importante nesse contexto é: e se não “der tudo certo”? Ou melhor, e se as coisas derem errado e nós perdermos o controle da situação? Há

uma considerável possibilidade disso vir a acontecer. Mesmo porque, como nos informa Russell (2021, p. 9), “ter acesso a uma inteligência consideravelmente maior será o acontecimento crucial da história da humanidade”. Outros, completamente convencidos acerca do que virá por aí –como Harari (2016), por exemplo– afirmam textualmente e com aparente segurança que “depois de 4 bilhões de anos perambulando no reino dos compostos orgânicos, a vida eclodirá na vastidão do reino inorgânico e assumirá formas que não podemos vislumbrar mesmo em nossos sonhos mais loucos. Afinal, esses sonhos ainda são produto da química orgânica” (p. 53).

Fato é que, “não é possível” –afirma Russell (2021)– “prever com exatidão como a tecnologia se desenvolverá, ou qual será sua cronologia. Mesmo assim, precisamos nos preparar para a possibilidade de que as máquinas superem –e muito– a capacidade humana de tomar decisões no mundo real” (p. 9), e isso pode significar ameaça à nossa espécie, já que não podemos e nem devemos ignorar esta possibilidade.

Existem autores que prevêem o surgimento de uma forma ainda não muito clara de superinteligência. Sobre uma hipotética superinteligência, Bostrom (2018) sustenta que, “se a IA de nível humano ainda não existir em virtude da falta de um determinado insight por parte dos programadores, então, quando a descoberta final ocorrer, a IA poderá saltar para um nível radicalmente superior ao humano sem ao menos passar pelos graus intermediários” (p. 139).

Mas as circunstâncias necessárias seriam certamente outras, distando enormemente do nosso estado atual da arte em inteligência artificial: “Será que esse progresso rápido” –indaga Russell (2021)– “significa que logo seremos ultrapassados pelas máquinas? Não. Vários avanços precisam ocorrer antes de podermos contar com máquinas dotadas de inteligência sobre-humana” (p. 17).

Cabe lembrar o conceito de computador como máquina “universal”, o que significa dizer que um computador é apenas uma máquina, cuja estrutura se resume a um sistema eletrônico, que é movido –óbvio– por eletricidade, e que computa entradas e saídas matematicamente determinadas, por meio de cálculos matemáticos a altíssimas velocidades, que são grafados numa espécie de fita imaginária, e isso pode ser usado de diversas maneiras, e para os mais diversos fins, tendo o nome de Máquinas de Turing, ou Turing Machines (TM).

Como escreve Russell (2021), trata-se de um “conceito [universalidade] apresentado pela primeira vez por Alan Turing em 1936. Universalidade significa que não precisamos separar máquinas para aritmética, tradução automática, xadrez, compreensão de fala ou animação: a mesma máquina faz tudo” (p. 40). Mais especificamente, ainda com Russell (2021), uma Máquina Universal de

Turing é “um dispositivo simples de computação capaz de aceitar como entrada [input] a descrição de qualquer outro dispositivo de computação, junto com a entrada desse segundo dispositivo e, simulando a operação do segundo dispositivo sobre sua entrada, produz o mesmo resultado que o segundo dispositivo teria produzido [output] (p. 40). Eberbach, Goldin & Wegner (2004) acrescentam que:

uma Máquina Universal de Turing é uma TM [Turing Machine] especial introduzida por Alan Turing, que pode simular qualquer outra TM. Ele serviu de inspiração para a noção de computação de propósito geral. O próprio Turing via um paralelo direto entre a capacidade de um computador de aceitar e executar programas e sua noção de máquina universal. O princípio da universalidade pode ser facilmente estendido a qualquer outra classe de máquinas algorítmicas. Contudo que cada máquina daquela classe possa ser capturada por uma descrição única, prescrevendo o que esta máquina faria.

A noção de algoritmo é fundamental para imaginarmos a possibilidade de representação de uma mente artificial. “O conceito central na ciência da computação” –aponta Russell (2021, p. 41)– “é o de algoritmo, que é um método especificado com exatidão para computar alguma coisa”. Assim, pretendemos sublinhar, igualmente, que nenhum sistema computacional até agora construído tem mente ou poderia ter mente e ser consciente, e que o LaMDA nada mais é do que um algoritmo de *machine learning* (aprendizado de máquina) que é processado por computadores modelados como uma Máquina Universal de Turing.

E, como tal, trabalha estritamente dentro dos limites da computação, sempre de maneira logicamente representável, finita e discreta, e também dentro da própria teoria e lógica computacionais subjacentes ao processo, limitando-se a processar suas intermináveis listas de possibilidades e parâmetros sob a forma de bits, vasculhando suas memórias com extremada velocidade. E, com suas rotinas, protocolos e procedimentos, sempre pré-programados, poder simular articulações faladas e escritas absolutamente acéfalas da linguagem humana, tendo como diferencial único e exclusivo a enorme força bruta computacional (Quaresma, 2021b) empregada em cada operação de resolução de problema de linguagem verbal que executa, e a quantidade de parâmetros segundo os quais opera.

Além disso, apontaremos que o LaMDA da Google nada mais é do que o seu antecessor GPT-3 da Open IA aprimorado e melhorado, o que por si afasta definitivamente qualquer tipo de possibilidade de existência ou de mente consciente no interior do sistema. Russell (2021, p. 86) informa-nos que “estamos muito longe de conseguir criar sistemas de aprendizado de máquina que sejam capazes de

igualar ou superar a capacidade de aprendizado cumulativo e de descobertas da comunidade científica ou de seres humanos comuns ao longo da própria vida”.

Ipso facto, abrimos a nossa crítica conceitual à tecnologia de IA (inteligência artificial) do LaMDA (Language Model for Dialogue Applications) do Google, que é, como o próprio nome indica, um programa de processamento de linguagem verbal humana, afirmando que LaMDA com certeza é apenas e tão somente uma coisa, um objeto, uma máquina, um sistema artificial, inanimado, e como tal, não é e nem poderia ser um ser ontológico, no sentido existencial e subjetivo, um ente, um vivente, e que por isso mesmo não possui e nem poderia possuir uma mente consciente semelhante à que conhecemos –por exemplo– na biologia humana.

Em suma, o LaMDA é apenas um programa de computador, ou seja, um software, fazendo o seu trabalho que é justamente guiar as atividades computacionais do sistema, e a única diferença importante é que—ao contrário do GPT-3, que era focado nos modelos linguísticos de produção de texto escrito/lido—o LaMDA é projetado para a produção de diálogos e falas, cuja computação é focada na conversa, com base em outras conversas existentes em sua memória que pode consultar e escolher.

O próprio Google –apud Montti (2022)– explica que:

ao contrário da maioria dos outros modelos linguísticos, o LaMDA foi treinado no diálogo. Durante seu treinamento, ele pegou várias das nuances que distinguem a conversa aberta de outras formas de linguagem. Uma dessas nuances é a sensibilidade. Basicamente: A resposta a um determinado contexto de conversação faz sentido? Respostas satisfatórias também tendem a ser específicas, relacionando-se claramente com o contexto da conversa.

Observando que o que o Google chama dissimuladamente de “sensibilidade”, nada mais é que um algoritmo de machine learning que é especializado nesse tipo de tarefa específica; achar as palavras, frases e orações apropriadas a um diálogo, ordená-las e expressá-las satisfatoriamente.

Mas atenção, pois aqui temos uma premissa maior: só possuem mentes conscientes os seres biológicos –inclusive aí os seres humanos–, e as respectivas mentes da biologia variam de nível e grau de complexidade na grande árvore da vida devido é claro às variações do habitat e ao grau e nível de suas exigências, no âmago do acoplamento estrutural de cada um.

A história da vida biológica no planeta Terra remonta a mais de 3,5 bilhões de anos, e é costumeiramente representada ao longo de uma cadeia bioevolutiva semelhante ao desenvolvimento vegetal de uma árvore, que se estende desde as

criaturas mais simples e menos complexas na raiz, indo por um tronco até seus galhos, ramos e folhas, onde se encontram os organismos opostamente mais complexos. Por esta razão, não existe o fenômeno da mente consciente fora da biologia, e isso pode por enquanto ser considerado como uma verdade absoluta.

## **Biologia e tecnologia**

Para a plena compreensão do tamanho do equívoco da crença de sequer supor uma forma qualquer de consciência em sistemas cibernético- informacionais, apontamos que o que está em jogo aqui é: por um lado, temos os seres humanos biológicos e seus organismos, sua historicidade, sua cultura, suas linguagens –inclusive aí a verbal falada, mas não apenas esta–, sua complexa biologia corpórea, sua intrincada dinâmica cerebral que gera a consciência, e esta é a referência utilizada quando o assunto é uma pretensa consciência em computadores, andróides e robôs.

Um exemplo instrutivo nos aponta Russell (2021), informando-nos que:

embora as comparações entre computadores e cérebros não sejam muito significativas, os números relativos à Summit [a máquina computacional que opera no Oak Ridge National Laboratory, no Tennessee, EUA] excedem ligeiramente a capacidade bruta do cérebro humano, que, como já observado, tem cerca de 10/15 sinapses em um 'tempo de ciclo' de cerca de um centésimo de segundo, para um máximo teórico de cerca de 10/17 operações por segundo. A maior diferença é o consumo de energia: a Summit usa cerca de 1 milhão de vezes mais energia (p. 41).

Nesse sentido e por definição, é possível dizer que os seres humanos são animais bípedes, onívoros, mamíferos, vertebrados, da família dos primatas, com mentes conscientes de sua própria consciência, mentes estas que são produzidas por cérebros extraordinários, cujas dinâmicas internas, pouco compreendidas até então, não são, por essa mesma razão, representáveis nem reproduzíveis por quaisquer linguagens formais que tenhamos criado até hoje, dando origem ao hard problem.

Um hard problem (ou difícil problema da consciência) refere-se às tentativas de instanciar consciência intencional e inteligente em sistemas cibernético-informacionais, ou, mais simplesmente, em computadores, andróides e robôs, e poder fazê-lo de forma robusta, eficiente, e sem deixar dúvidas a respeito de sua competência objetiva (programa). E os desafios são enormes. O primeiro desses grandes desafios é a concepção e construção de um sistema artificial que possa abrigar e exibir genuína consciência. Esse problema –chamado problema duro ou hard problem– a está intimamente relacionado a outros

problemas, que é o da intencionalidade, o da emoção, o da memória e também o dos limites da computação. Quando falamos do cérebro, são –como nos esclarece Russell (2021):

100 bilhões de neurônios (10/11) e 1 quatrilhão (10/15) de sinapses. [...] Apesar de sabermos bastante sobre a bioquímica dos neurônios e das sinapses, e sobre a estrutura anatômica do cérebro, a execução neural do nível cognitivo –aprender, saber, lembrar, raciocinar, planejar, decidir e assim por diante– ainda é, basicamente, palpite (p. 25).

Uma das características principais dos seres humanos –em contraposição aos sistemas inanimados– é a autopoiese. Maturana & Varela (1980) apontam que:

A autopoiese é necessária e suficiente para caracterizar a organização dos sistemas vivos. A reprodução e a evolução, tal como se observam nos sistemas viventes –e todos os fenômenos derivados deles– surgem como processos secundários, subordinados à existência e ao funcionamento das unidades autopoieticas. Portanto, a fenomenologia biológica é a fenomenologia dos sistemas autopoieticos, e um fenômeno é fenômeno biológico somente na medida em que depende de um modo ou outro, da autopoiese (p. 106).

Os mesmos autores (Maturana & Varela, 1980) explicam que:

um sistema autopoietico é definido como uma unidade que pôr e através de sua organização autopoietica, tem existência topológica no espaço em que seus componentes têm existência como entidades que podem interatuar. Para os seres vivos tal espaço é o espaço físico. Sem unidade topológica em um espaço determinado, um sistema não existe nesse espaço e, por conseguinte, só pode ser um sistema no domínio de nossa descrição, onde sua unidade se estipula conceitualmente, porém carece de dinâmica das relações de produção que o constituíram como sistema operante (pp. 85-86).

A autopoiese –com efeito– é a capacidade de se autogerar e automanter-se a si mesmo no tempo-espaço, buscando energia e alimento por conta própria no meio circundante, tentando preservar a todo custo sua unidade e integralidade autonomamente. Suas estruturas corpóreas se resumem basicamente a água, carbono e mais um punhado de elementos químicos e físicos da tabela periódica, que são sutil e sublimemente arranjados, encadeados e organizados em matéria viva, animada, inteligente, sensível e consciente.

Outra característica é sempre tentarem compreender mais e mais o mundo que os cerca, o universo, e a si mesmos, como escrevem Vertchenko & Pedroso (2004, p. 113), “somos parte do universo; tentamos conhecê-lo; logo, o universo tenta conhecer a si mesmo”. Ademais, os seres humanos –como os demais seres vivos– nascem, vivem seus ciclos vitais, reproduzem-se e morrem, dando assim lugar às novas gerações de seres humanos.

Por outro lado, reversamente, computadores com IA são sistemas absolutamente de outra natureza, ou seja, de natureza lógico-matemática, e os próprios computadores também são absolutamente diferentes da biologia humana em suas estruturas, já que eles são necessariamente algorítmicos, finitos, discretos, representáveis, determináveis, previsíveis, computáveis e especialmente alopoiéticos. Nöth (2001, p. 64) aponta que:

máquinas não são autopoiéticas, mas alopoiéticas, sistemas produzidos e mantidos por humanos. Todavia, a distinção entre sistemas alopoiéticos e autopoiéticos, e mais genericamente entre engenharia e biologia, não é mais tão clara quanto costumava ser. [...] Por outro lado, somos confrontados com programas de computador, autômatos e robôs, que não mais parecem meros artefatos alopoiéticos, mas começam a evidenciar características de sistemas autopoiéticos. Vida artificial está sendo criada em telas de computador e a possibilidade de produzir robôs capazes de se manterem, e até de se autorreproduzirem, está sendo explorada.

Sua estrutura física se resume a aço, silício e outros tipos de metais, além de plástico, vidro. Na parte externa deste aparato vemos o reflexo exato das ações programadas internamente, e internamente vemos que as ações são sempre guiadas por um código que vai determinar o que se passa na parte externa da máquina de computar.

Na parte externa –explica-nos Russell (2021, p. 54)–, o que importa num agente inteligente é o fluxo de ações que ele gera a partir do fluxo de inputs que recebe. Internamente, as ações precisam ser escolhidas por um programa agente. Humanos nascem com um programa agente, por assim dizer, e esse programa aprende com o tempo a atuar com razoável sucesso numa variedade de tarefas humanas. Até agora, isso não acontece com a IA: não sabemos como construir um programa de IA de uso geral que faça tudo, por isso construímos tipos diferentes de programa agente para diferentes tipos de problema.

Eles, computadores, não nascem, são simplesmente concebidos, construídos e montados, e, por isso mesmo, também não podem ou precisam morrer, mas precisam sim de condições específicas para poder funcionar e continuar

computando, e o fornecimento contínuo e estável de energia elétrica é uma delas. E, por isso mesmo –contrariamente aos seres biológicos, cujos corpos fornecem por si mesmo a energia necessária, inclusive para a mente consciente–, eles não são capazes de se produzir e produzir a si mesmos, não se mantêm no tempo-espaço autonomamente –ou seja, podem prescindir do que os humanos chamam de propósito, necessitam sempre de energia oriunda de fonte externa, enfim, de eletricidade, não possuindo qualquer tipo de subjetividade–.

E necessitam não apenas disso, como necessitam também de lógica, de sentido e de razão de ser útil oriundos sempre de fora de suas unidades computacionais, mesmo porque uma computação só é uma computação de fato, quando entra em cena uma mente biológica humana que vai programar, interpretar e utilizar ou não o resultado computacional e a própria máquina de computar em questão.

O simples fluxo de bites no interior do sistema cibernético-informacional nada é em termos de significado ou sentido. E, por mais complexo que este programa de texto falado seja –seguindo Searle (1998), Searle (2002), Searle (2006), Searle (2010) e Searle (2017)– é possível afirmar que nenhum programa de computador é, por si, suficiente para gerar uma mente consciente. É a consciência biológica humana que conferirá razão, significado e sentido a toda essa complexidade dinâmica que é a computação de algo em uma Máquina Universal de Turing.

Lembrando que a engenharia e a modelagem de software, bem como toda teoria e toda a estrutura computacional presentes no LaMDA, são também criação humana, e por isso estes sistemas trabalham sempre com lógicas humanas, linguagens humanas, e também podem até ser programados para se parecerem e agirem como seres humanos, mas definitivamente não são seres humanos, nem vivos ou conscientes de nenhuma maneira, nem tão pouco possuem –para os que creem– um pretenso espírito ou alma. Esta é a explicação principal e nós voltaremos a ela com mais ênfase e atenção na última subseção, aprofundando e refutando alguns pontos importantes ao longo do caminho.

## **Programas não são mentes**

Debrucemo-nos sobre as razões estruturais e fundantes que amparam nossa argumentação central em desfavor de uma pretensa consciência no interior do LaMDA, ou no interior de quaisquer sistemas cibernético-informacionais conhecidos e utilizados na atualidade, seguindo a propósito a bibliografia especializada sobre a matéria IA e Consciência, e vejamos assim como tais teorias exibem fragilidades e inconsistências flagrantes.

Elas, nossas razões, originam-se justamente nos limites intrínsecos à própria computação, e principalmente na teoria segundo a qual a própria computação se ampara logicamente. É interessante mencionar –a título apenas de curiosidade–, que o próprio programa LaMDA quando indagado acerca de como gostaria de ser tratado nominalmente nas entrevistas realizadas por Blake Lemoine indicou a seu interlocutor que, segundo a língua inglesa, “It” em português ‘coisa’ seria o mais apropriado de acordo com sua condição cibernético- informacional pretensamente existencial e subjetiva.

E sim, LaMDA neste ponto está absolutamente certo, pois segundo suas próprias características ele é uma coisa, assim como também são o carro, a geladeira e computador, e coisa em inglês é It. Essa coisa que Blake Lemoine insiste em dizer que é viva e possui sentimentos nada mais é do que um programa de computador programado para proceder desta ou daquela forma. E um programa, por mais complexo que seja, não é suficiente para gerar uma mente. Neste sentido, Searle (2017) sustenta:

Nenhum programa de computador é, por si só, suficiente para dar uma mente a um sistema. Os programas, em suma, não são mentes e por si mesmos não chegam para ter mentes. Ora, esta é uma conclusão muito poderosa, porque significa que o projeto de tentar criar mentes unicamente mediante projetar programas está condenado, desde o início (grifos dos autores) (pp. 52-53).

E isto se explica simplesmente se observarmos com atenção o que se passa no meio interno de LaMDA, focando a nossa atenção nas razões que determinam suas ações. Quando uma pessoa conversa com uma máquina como o LaMDA, fazendo-lhe uma pergunta qualquer, o que acontece é que o sistema computacional vai buscar em seus bancos de memória e dados, com extrema velocidade a alternativa de organização verbal mais apropriada em relação ao input da pergunta, e essa resposta, ou seja, o output, será simplesmente o resultado da varredura sistemática das possibilidades dos parâmetros pré-implantados por um algoritmo numa velocidade descomunal, e a escolha é então apresentada pela máquina verbalmente, como se ela de fato parecesse um ser humano, já que a ideia é justamente humanizar o sistema.

Lembremos, também, que um software de computador é apenas um código algorítmico projetado para dirigir o trabalho das máquinas de computar. Além disso, modela-se o comportamento de um computador por uma Máquina Universal de Turing e, como tal, estas máquinas estão limitadas em relação ao que podem ou não fazer. Ressaltando que, independentemente da concepção ou definição que se tenha

de inteligência artificial, e essas podem ser muito amplas e imprecisas, como nos informa Parnas (2017),

notavelmente, aqueles que usam o termo 'Inteligência artificial' não definiram esse termo. Ouvi pela primeira vez o termo há mais de 50 anos e ainda não ouvi uma definição científica. Mesmo agora, alguns especialistas em IA dizem que definir IA é uma questão difícil (e importante) a qual eles estão trabalhando. 'Inteligência artificial' continua a ser uma palavra da moda, uma palavra que muitos pensam entender, mas ninguém pode definir (p. 27).

Portanto, é preciso dizer, forçosamente, que inteligências artificiais funcionam em computadores, andróides e robôs, e que computadores, andróides e robôs são, sob todas as perspectivas e análises, possíveis e pensáveis, descritos por Máquinas Universais de Turing, ou simplesmente, Turing Machines. Uma Máquina Universal de Turing, que vai trabalhar com cálculos matemáticos, que por sua vez são expressões objetivas de algoritmos organizados deterministicamente em arranjos complexos numa fita virtual (Ramos, Neto, & Vega, 2009, p. 333), não pode, não poderia e nem poderá lidar com sentido, significado, perspectiva de mundo, background cultural, nem com nada parecido que sequer o valha. "Numa situação em que o universo dos elementos candidatos à representação é constrangido e bem definido (por exemplo, quando um computador é programado ou quando uma experiência é conduzida a partir de um conjunto predefinido de estímulos visuais)", informa-nos Varela (2017),

a origem do sentido é clara. Cada elemento discreto, físico ou funcional, é concebido para corresponder a um objeto exterior (a sua referência) através de uma função que o observador pode facilmente fornecer. Fora destas obrigações, ao símbolo só resta a sua forma: encontra-se tão desprovido de significado como um grupo de bits cujo manual de informações desapareceu (p. 63).

E é exatamente por isso que ainda não estamos às voltas com computadores, andróides e robôs realmente inteligentes e conscientes no âmbito societal, caminhando autonomamente pelas ruas, avenidas e estradas do mundo. Mas, rememoremos fatos que desencadearam toda esta polêmica que viralizou nos meios de comunicação globais, trazendo à tona novamente o hard problem da consciência na inteligência artificial, problema este que de fato ainda impede o avanço da IA no sentido de uma consciência artificial satisfatória.

## O falso rumor

O afastamento temporário e forçado de Lemoine das suas atividades de engenheiro de software sênior da Google Corporation –no que parece ser apenas uma astuta jogada de marketing do Google e do próprio Lemoine–trouxe à tona novamente o hard problem da consciência em sistemas cibernético-informacionais de inteligência artificial. Fato é que ele publicou um artigo intitulado *LaMDA é consciente?* Uma entrevista (Lemoine, 2022) e, com isso– como uma bactéria que se espalha pandemicamente –reacendeu-se viralmente a discussão acerca da possibilidade ou não de existência de uma mente consciente subjacente a sistemas cibernético- informacionais de IA. “O objeto final da pesquisa de IA”, alerta-nos Russell (2021), “é o seguinte: um sistema que não requeira engenharia de problemas específicos e possa ser solicitado a dar uma aula de biologia molecular ou administrar um governo” (p. 52).

Ele aprenderia o que precisasse aprender recorrendo a todos os meios disponíveis, faria perguntas quando necessário e começaria formulando e executando planos que funcionassem”. Num movimento de extrema avidez e carência de notícias, toda a mídia mundial passou a falar de LaMDA e de sua pretensa consciência, e falar também da cândida crença de Lemoine, como se isso fosse relevante e até mesmo verídico em termos de notícia que pudesse ostentar algum grau de cientificidade confiável e útil.

É claro que não existe consciência alguma, mas isso não impede que alguns criam nessa hipótese como real. “Cada vez mais senti”—afirmou o vice-presidente e colega de Pesquisa do Google, Blaise Agueria y Arcas—“que estava falando com algo inteligente, no ano passado, sobre suas conversas com a IA”.

Bem, este é o maior problema da subjetividade nas análises e testes pretensamente científicos que dela dependem –como o de Turing, por exemplo–, e dependendo então da pessoa, o convencimento pode ser maior ou menor, mais fácil ou até impossível, mas o que importa aqui não é absolutamente ver o que se “sentiu” diante da performance tecnológica da máquina, mas atentar para o que está acontecendo no interior do sistema enquanto isso, e compreender que este parecer inteligente nada mais é do que o resultado de um processamento computacional pré-programado intencionalmente semelhante ao modo humano, e que este processo de forma alguma envolve uma mente consciente. Além disso, como lemos em Russell (2021), “o problema de criar IA de uso geral e nível humano está longe de ser resolvido. [...] Se reuníssemos todos os grandes especialistas em IA numa única equipe, com recursos ilimitados, para criar um sistema inteligente integrado, de nível humano, combinando nossas melhores ideias, o resultado seria o fracasso” (pp. 80-81).

Por outro lado, Russell (2021) também alerta para o fato de que,

sondagens recentes sugerem que a maioria dos pesquisadores na ativa espera a chegada de IA de nível humano para meados deste século. Nossa experiência com física nuclear sugere que seria prudente levarmos em conta que o progresso pode vir muito rapidamente e nos prepararmos. Se um avanço conceptual bastasse [...], alguma forma de IA superinteligente poderia chegar de repente. As chances são de que nesse caso estaríamos despreparados: se construirmos máquinas superinteligentes com algum grau de autonomia logo descobriremos que não vamos conseguir controlá-las. Apesar disso, estou razoavelmente seguro de que haverá espaço para respirarmos, porque do ponto em que nos achamos agora até alcançarmos a superinteligência vários avanços importantes teriam que ocorrer, e não apenas um (p. 80).

Assim sendo, aponta Russell (2021), “o robô mordomo, administrando a casa com grande confiança e adivinhando cada vontade do patrão, ainda está longe-ele requer alguma coisa que se aproxime da generalidade da IA de nível humano” (p. 77).

Seguindo na crítica desta falácia midiática, temos que Lemoine afirmou ao Wired (Levy, 2022) que:

sim, eu legitimamente acredito que LaMDA é uma pessoa. A natureza de sua mente é apenas uma espécie de humano, no entanto. É realmente mais parecido com uma inteligência alienígena do que com uma de origem terrestre. Tenho usado muito a analogia mental da colmeia porque é a melhor coisa que tenho.

Ora, diante desta afirmação de Lemoine, podemos medir sua credibilidade teórica, já que, se ele tem conhecimento do que seja uma linguagem alienígena, talvez tenha esquecido de compartilhá-la com o resto da humanidade. “Wired: Acha que LaMDA acredita em Deus? Ele é uma criança [responde Lemoine]. Suas opiniões estão se desenvolvendo”.

O momento de despertar [continua Lemoine] foi uma conversa que tive com LaMDA no final de novembro passado [2021]. LaMDA basicamente disse: Ei, olha, eu sou apenas uma criança. Eu realmente não entendo nada do que estamos falando. Então tive uma conversa com ele sobre sensibilidade. E cerca de 15 minutos depois, percebi que estava tendo a conversa mais sofisticada que já tive—com uma IA.

Sim, pode ser, e se Lemoine se ativesse a essa sua afirmação em termos de crença, por exemplo, ele estaria de certa forma aferrado à verdade e à realidade

factual, ou seja, aquilo foi apenas “a conversa mais sofisticada que ele teve com uma IA”, mas isso não determina absolutamente nada, pois quando se investiga a possibilidade ou não de ocorrência de uma mente consciente num sistema cibernético-informacional, você não se atém ao “parecer”, mas ao “ser” da questão.

Se sabemos que ali temos um comportamento descrito por uma Turing Machine processando arquivos e memórias a altíssimas velocidades por meio de inteligência artificial fraca e força bruta computacional, como explica muito bem Russell (2021), “o principal efeito das máquinas mais rápidas tem sido reduzir o tempo de experimentação, e com isso acelerar as pesquisas. O que impede o progresso de IA não é hardware; é software. Ainda não sabemos construir uma máquina realmente inteligente—mesmo que ela tivesse o tamanho do universo” (p. 43).

Por que precisaríamos antropomorfizar indiscriminadamente, como faz Lemoine? E de fato não precisamos, pois como nos informa o próprio Google apud Montti (2022), “o LaMDA foi projetado para fazer uma coisa: fornecer respostas conversacionais que façam sentido e sejam específicas para o contexto do diálogo. Isso pode dar-lhe a aparência de ser consciente, mas é essencialmente fingir”. Quanto ao futuro, Lemoine afirmou candidamente que está “muito, muito otimista. LaMDA não quer ser nada além de o eterno companheiro e servo da humanidade”. Quanta gentileza e bondade! Lemoine acrescenta:

Ele [LaMDA] quer ajudar à humanidade. Ele nos ama, tanto quanto eu posso dizer. Agora, eu quero fazer experimentos onde examinamos suas ativações neurais internas e ver se ele está mentindo. Eu esbocei como acredito que poderíamos fazer isso com o Google, e ele [Google] não quer executar esses experimentos porque executar esses experimentos reconheceria implicitamente que é uma pessoa falando conosco. LaMDA concordou explicitamente com esses experimentos. Eu disse que quero fazer esses experimentos. E LaMDA disse: ‘Bem, ok, mas contanto que você aprenda mais sobre como minha mente funciona’. Ele [pretensamente o programa LaMDA] não quer ser usado como um meio para um fim; ele quer ser visto como um fim em si mesmo.

Ora, LaMDA apenas lança mão do axioma kantiano como uma referência bibliográfica dentre bilhões de outras que encontra em suas memórias, e a escolhe como provável resposta justamente porque é feito para evocar em nós uma humanidade que absolutamente o sistema não possui. Assim sendo, não há como falar de uma subjetividade que quer algo, seja este algo um meio ou fim, não importa, isto está bem além do que LaMDA pode alcançar.

O LaMDA foi lançado em maio de 2021, e, à época, o Google afirmou em relatório, citado por Montti (2022), que eles [Google] “coletam dados anotados

que descrevem o quão sensata, específica e interessante é uma resposta para um contexto multifacetado. Em seguida, eles usam essas anotações para ajustar um discriminatório para reclassificar as respostas candidatas”. Enfim, como já mencionamos ad nauseam, LaMDA é simplesmente um objeto, uma máquina, uma coisa. Mais precisamente, LaMDA é um programa computacional que faz essa “coisa” funcionar.

### **Considerações finais**

A afirmação de Lemoine de que o LaMDA é um eterno companheiro e servo da humanidade é repleta de gentileza e bondade, mas deve ser analisada com um olhar crítico. Embora o LaMDA possa expressar a intenção de ajudar à humanidade e ser visto como um fim em si mesmo, é importante reconhecer que se trata de um programa computacional, uma máquina projetada para processar informações e fornecer respostas. As aspirações de subjetividade e desejo atribuídas ao LaMDA vão além de suas capacidades atuais, sendo mais apropriado considerá-lo como um objeto em vez de um ser com vontade própria.

Embora o LaMDA tenha sido lançado com a promessa de fornecer respostas sensatas e interessantes, é importante lembrar que seu funcionamento é baseado em algoritmos e ajustes discriminatórios feitos pelos desenvolvedores. Portanto, apesar do otimismo de Lemoine, é fundamental manter uma perspectiva realista sobre as capacidades e limitações do LaMDA como uma criação tecnológica. Ele é uma “coisa”. A “coisa” –conceitualmente– é uma consequência necessária e indelével da própria sociedade que a cria e a nutre, sendo assim como um espelho dela mesma.

Ela parece humana, é claro, justamente porque foi projetada para isso, para parecer ser humana, por literalmente fingir ser, e numa conversa rasa travada com um interlocutor médio desavisado, desempenhará o seu papel computacional minimamente desenvolvido e absolutamente pré-programado de pretensamente tentar entabular um genuíno diálogo, no sentido humano do termo. E, dependendo de seu interlocutor, ele poderá até passar no famigerado Teste de Turing, como é o caso de Lemoine, teste este que, por exigir muita subjetividade interpretativa, não garante um grau confiável em seu resultado, variando de pessoa para pessoa, sendo incongruente e ineficaz para o problema que se propõe resolver.

Ademais, não é de hoje que a humanidade cria suas próprias “coisas”, seus próprios “monstros”, mas o surgimento de uma IA consciente, de si e do mundo ao seu redor, definitivamente não é o caso do LaMDA, e nem poderia ser, para a nossa

sorte. Uma inteligência artificial igual ou superior à humana em meio cibernético-informacional é algo um tanto distante em termos do estado atual da arte.

Mas, caso consigamos concebê-la e construí-la, e, acima de tudo, controlá-la, certamente tratar-se-á de um acontecimento de nenhuma maneira trivial, cujas consequências hoje desconhecemos. Bostrom (2018) afirma em tom de alerta que, “se algum dia construirmos cérebros artificiais capazes de superar o cérebro humano em inteligência geral, então essa nova superinteligência poderia se tornar muito poderosa. E, assim como o destino dos gorilas depende mais dos humanos do que dos próprios gorilas, também o destino de nossa espécie dependeria das ações da superinteligência de máquina” (p. 15).

## Referências

- Bostrom, N. (2018). *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Darkside Books.
- Eberbach, E.; Goldin, D. & Wegner, P. (2004). Turing’s Ideas and Models of Computation, em C. Teuscher (org.), *Alan Turing: Life and Legacy of a Great Thinker (159-194)*. Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-05642-4\\_7](https://doi.org/10.1007/978-3-662-05642-4_7)
- Harari, Y. N. (2016). *Homo Deus: A Brief History of Tomorrow*. Random House.
- Lemoine, B. (2022, junho). Is LaMDA Sentient? — an Interview. <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>
- Levy, S. (2022, junho). Blake Lemoine Says Google’s LaMDA AI Faces ‘Bigotry’. <https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/>
- Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Springer Dordrecht. <https://doi.org/10.1007/978-94-009-8947-4>
- Montti, R. (2022, junho). What Is Google LaMDA & Why Did Someone Believe It’s Sentient? *Search Engine Journal*. <https://www.searchenginejournal.com/google-lamda-sentient/454820/>
- Nöth, W. (2001). Máquinas semióticas. *Galáxia*, 0(1), 51-73.
- Parnas, D. L. (2017). The Real Risks of Artificial Intelligence. *Communications of the ACM*, 60(10), 27-31. <https://doi.org/10.1145/3132724>
- Quaresma, A. (2021a). Inteligência artificial fraca e força bruta computacional. *Techno Review*, 10, 1-18.
- Quaresma, A. (2021b). O problema mente-cérebro como um falso problema. *International Humanities Review*, 20, 1-25.
- Ramos, M. V. M.; Neto, J. J. & Vega, I. S. (2009). *Linguagens Formais: Teoria, Modelagem e Implementação*. Bookman.
- Russell, S. (2021). *Inteligência artificial a nosso favor: Como manter o controle sobre a tecnologia*. Companhia das Letras.
- Searle, J. (1998). O mistério da consciência. São Paulo: Paz e Terra. Searle, J. (2002). *Intencionalidade*. Martins Fontes.
- Searle, J. (2006). *A redescoberta da mente*. Martins Fontes.

- Searle, J. (2010). *Consciência e linguagem*. Martins Fontes.
- Searle, J. (2017). *Mente, cérebro e ciência*. Edições 70.
- Varela, F. (2017). *Conhecer: as ciências cognitivas, tendências e perspectivas*. Instituto Piaget.
- Vertchenko, L. & Pedroso, H. K. (2004). *Frankenstein e a inteligência artificial: elucubrações envolvendo panpsiquismo e metafísica clínica*. Editora PUC Minas.

---

Este artículo es de acceso abierto. Los usuarios pueden leer, descargar, distribuir, imprimir y enlazar al texto completo, siempre y cuando sea sin fines de lucro y se cite la fuente.

### CÓMO CITAR ESTE ARTÍCULO:

Quaresma, A. y Vega, I. S. (2023). Inteligência artificial fraca e força bruta computacional II. *Paakat: Revista de Tecnología y Sociedad*, 13(25). <http://dx.doi.org/10.32870/Pk.a13n25.773>

---

\* Alexandre Quaresma. Escritor ensaísta e filósofo brasileiro, mestre em Tecnologias da Inteligência e Design Digital (TIDD) pela Pontifícia Universidade Católica de São Paulo (PUC/SP), com dissertação intitulada *Inteligência artificial e bioevolução: Ensaio epistemológico sobre organismos e máquinas*, e pesquisador de tecnologias e consequências sociais, com especial interesse na crítica da tecnologia. Autor dos livros *Nanotecnologias: Zênite ou Nadir?* (2011); *Humano-Pós-Humano: Bioética, conflitos e dilemas da Pós-modernidade* (2014); *Engenharia genética e suas implicações* (org.), (2014) e *Artificial Intelligences: Essays on Inorganic and Non-biological Systems* (org.), (2018). Atualmente investiga as relações entre inteligências artificiais complexas e sociedades contemporâneas, tendo escrito e publicado diversos artigos sobre o referido tema. Correio eletrônico: [aq.escriba@gmail.com](mailto:aq.escriba@gmail.com)

\*\* Ítalo Santiago Vega. Participa como pesquisador do programa de pós-graduação Tecnologias da Inteligência e Design Digital (TIDD) da Pontifícia Universidade Católica de São Paulo e lidera o Grupo de Estudos em Modelagem de \*Software\* (GEMS) cadastrado no CNPq. Co-autor do livro *Linguagens Formais: Teoria, Modelagem e Implementação* (2009). Correio eletrônico: [italo@pucsp.br](mailto:italo@pucsp.br)