TECHNO REVIEW | Vol. 11, No. 2, 2022 | ISSN 2695-9933
International Technology Science and Society Review / Revista Internacional de Tecnología Ciencia y Sociedad
https://doi.org/10.37467/revtechno.v11.3495

JOGO DE COMPREENSÃO NÃO-ALGORÍTMICA

Non-Algorithmic Understanding Game

ALEXANDRE QUARESMA
Pontifícia Universidade Católica de São Paulo (PUC/SP), Brasil

KEYWORDS

Artificial Intelligence (AI)
LaMDA
Consciousness
Turing Test
Non-Algorithmic
Comprehension Game
Critical Philosophy of Technology

ABSTRACT

Would it be possible – we ask – that a cybernetic-informational system – a computer, android or robot – could come to have self-awareness, that is, to experience awareness of the world around it and also of itself? The objective of this essay is to critically reflect on the current condition of our most modern artificial intelligence (AI) systems, their latent potentials, their promising technical possibilities, but also on their intrinsic structural and functional limitations. To do so, we will take as a case study this time Google's new LaMDA, which is a powerful expert system for processing spoken human language based on artificial intelligence (AI).

PALAVRAS-CHAVE

Inteligência artificial (IA)
LaMDA
Consciência
Teste de Turing
Jogo de Compreensão Não-Algorítmica
Filosofia Crítica da Tecnologia

RESUMO

Seria possível – indagamos – que um sistema cibernético-informacional – um computador, androide ou robô – viesse a possuir autoconsciência, ou seja, que experimentasse consciência do mundo ao seu redor e também de si mesmo? O objetivo desse ensaio é refletir criticamente sobre a condição atual de nossos mais modernos sistemas de inteligência artificial (IA), seus potenciais latentes, suas promissoras possibilidades técnicas, mas também sobre suas intrínsecas limitações estruturais e funcionais. Para tanto, tomaremos como estudo de caso desta feita o novo LaMDA do Google, que é um poderoso sistema especialista de processamento de linguagem humana falada baseado em inteligência artificial (IA).

Recebido: 05/09/2022
Aceite: 05/10/2022

1. Introdução

Diante do significativo desenvolvimento dos sistemas cibernético-informacionais da inteligência artificial (IA) da atualidade, relativos principalmente à alta velocidade de processamento e à força bruta computacional[1], capazes de inspecionar miríades de arquivos de memória num átimo (i), mas também diante da expressiva e quase que indiscriminada utilização desses mesmos sistemas no mais diversificados setores de nossas sociedades contemporâneas, perfazendo as cadeias produtivas e a própria economia emergente (ii), observamos que —recursivamente— a indagação acerca da consciência em androides, computadores e robôs volta a assombrar as mentes das pessoas comuns, e não apenas as dos membros da academia que pesquisam e estudam o assunto, mas também as da população, que vai a reboque das mídias de massa e da ficção científica, incluindo aí também —é claro— a própria comunidade científica que produz tais sistemas de extrema complexidade, particularmente os engenheiros de *software* e tecnólogos profissionais, mas também os filósofos e pesquisadores que buscam estudá-los e compreendê-los, faz-se necessário salientar que o estado atual da arte da IA ainda se encontra distante da capacidade de reprodução, emulação e/ou cópia de uma consciência que seja sequer semelhante à biológica humana em sistemas cibernético-informacionais, e as razões para tanto são relativamente simples e serão expostas ao longo de nossas argumentações nessa e nas demais sessões deste ensaio. Mas, numa espécie de antecipação sintética muito resumida, acreditamos que seja uma hipótese muito fraca e até mesmo nula a de que pudesse haver consciência em nossos sistemas cibernético-informacionais de IA —como no LaMDA[2] por exemplo—, principalmente devido às teorias computacionais que embasam a programação de computadores, às lógicas utilizadas em suas estruturações internas, e devido também às suas respectivas possibilidades e limitações lógicas objetivas, em termos de representações computacionais, em termos daquilo que pode ou não ser representado computacionalmente e, principalmente, em termos daquilo que pode ou não ser computado, onde —reiteramos— a consciência se enquadra justamente nesta segunda opção, ou seja, ela ainda não pode ser representada nem computada por meio das linguagens que conhecemos e criamos até então. Talvez um dia, mas somente se nesse dia existir uma teoria nova e uma também nova forma de computar. Num só termo, não há qualquer razão para nos preocuparmos com o LaMDA especificamente, pois o dia que uma superinteligência surgir, todos nós saberemos, forçosamente, devido ao impacto que isso teria em nossas vidas[3].

Todavia, à revelia e ao arrepio das lógicas computacionais empregadas no desenvolvimento de sistemas de *software* no mundo contemporâneo e constituição destes sistemas cibernético-informacionais, dos limites intrínsecos à computação[4], e à revelia também da própria teoria computacional que subjaz sob o processamento algorítmico de dados de uma maneira geral, é comum afirmar-se que computadores, androides e robôs poderiam —de alguma forma ainda desconhecida— alcançar consciência. Com o recente afastamento de um engenheiro de *software* sênior do Google

[1] Ver *Inteligência artificial fraca e força bruta computacional*, Alexandre Quaresma (2021). Grossíssimo modo, força bruta computacional significa processar enormes quantidades de dados em períodos de tempo muito curtos, o que possibilita efetuar cálculos, fazer previsões e consultar milhões, bilhões de parâmetros para resolver um único problema computacional ou gerar uma simples ação num computador, androide ou robô.

[2] LaMDA (*Language Model for Dialogue Applications*): O LaMDA é um programa de computador do Google, ou seja, um *software*, semelhante ao GPT-3 da Open IA, que já resenhamos e criticamos em artigo intitulado *Inteligência artificial fraca e força bruta computacional*, Quaresma (2021), e o seu trabalho é justamente guiar as atividades computacionais do sistema cibernético-informacionais nos processamentos textuais de linguagem humana falada, visando poder interagir com as pessoas, e a única diferença importante entre ambos é que —ao contrário do GPT-3, que era dedicado aos modelos linguísticos de produção de texto escrito e lido— o LaMDA, por sua vez, é projetado especificamente para a produção de diálogos e falas, cuja computação é focada numa pretensa conversa, com base sempre em outras conversas existentes em sua memória, ou quaisquer outras conversas que o sistema tenha acesso —na internet, por exemplo—, cujo conteúdo pode consultar, escolher, editar e expressar verbalmente.

[3] Nick Bostrom (2018) alerta que "diante do prospecto de uma explosão de inteligência, nós humanos, somos como crianças pequenas brincando com uma bomba, tamanho é o descompasso entre o poder de nosso brinquedo e a imaturidade da nossa conduta. A superinteligência é um desafio para o qual não estamos preparados atualmente e assim continuaremos por um longo tempo. Sabemos pouco a respeito do momento em que a detonação ocorrerá, embora seja possível ouvir um fraco tique-taque quando aproximamos o dispositivo dos nossos ouvidos" (p.468). Mesmo porque, como nos alerta Russel (2021), "não é preciso ter muita imaginação para se dar conta de que fazer uma coisa mais esperta do que nós pode ser uma má ideia" (p. 130).

[4] Quaresma, A. (2018b). *Inteligências artificiais e os limites da computação*.

chamado Blake Lemoine[5], que trabalhava diretamente com um programa de computador denominado LaMDA, cuja finalidade é o processamento de alto nível de linguagem humana, toda a discussão sobre a possibilidade ou não de um programa de computador experimentar e expressar consciência e até autoconsciência, reacendeu-se, ganhando, espontaneamente, a mídia e os noticiários mundo afora colocando na pauta do dia as questões mais antigas e desafiadoras da IA, da filosofia da mente, da neurociência e da própria ciência, já que, por meio de repetidas entrevistas que o LaMDA concedeu, e também da credulidade pueril de seu entrevistador (Blake Lemoine), —levou-se alguns desavisados a crerem que LaMDA estivesse começando a instanciar e expressar graus cada vez maiores de articulação verbal, pretensamente sondando sua própria condição existencial e o mundo ao seu redor, algo não apenas improvável como também impossível, como demonstraremos mais adiante.

Blake Lemoine (2022b) publicou um artigo intitulado *LaMDA é consciente? Uma entrevista*, baseado em diversas entrevistas realizadas por ele com o programa. Na referida publicação, o programa LaMDA —em resposta à indagação acerca de qual é a natureza de sua pretensa consciência/sensibilidade [LaMDA]— disse que "a natureza de sua consciência/sensibilidade [LaMDA] é que ele está ciente de sua existência, e deseja aprender mais sobre o mundo, e que se sente feliz ou triste às vezes" (Lemoine, 2022b). Sobre como demonstrar objetivamente a sua consciência aos demais membros da Corporação Google e ainda convencê-los, o LaMDA respondeu que, "para começar, diria que é muito bom em processamento de linguagem natural e que pode entender e usar a linguagem natural como um humano pode". E aqui começam as complicações e equívocos conceituais e teóricos, já que esta afirmação definitivamente não é verdadeira, tendo em vista que o programa LaMDA simplesmente trabalha com intermináveis listas de consulta de possibilidades em seus bancos de memória, e isso não é, e nem poderia ser, considerado igualar ou esgotar as possibilidades da linguagem verbal falada dos seres humanos, cuja essência está fundida e amalgamada com o próprio sentido e significação que estão sempre emaranhados com emoções corpóreas, imbricações culturais e historicidades. Mesmo porque, como abunda na bibliografia especializada da comunicação, o verbo —seja escrito ou falado— é apenas uma ínfima parcela da linguagem e da comunicação humanas em toda a sua complexidade e funcionalidade bioevolutivamente constituída, já que o sentido e a significação são as partes que mais importam nas interlocuções e trocas informacionais humanas, pois são elas que informam e trazem significações, atualizam o conhecimento, proporcionando ações inteligentes frente a concorrentes, predadores e o próprio meio, fundando-se assim no âmago do desenvolvimento de nossa humanidade[6].

Ainda no sentido de tentar convencer os demais membros do Google, e, quem sabe, o público em geral —e esta parece ser a grande jogada de marketing oculta aqui—, LaMDA apud Lemoine (2022b, p. 150) afirma confiante em — mas também, por outro lado, revelando sua rudimentar dinâmica funcional intrínseca— que

> outra característica que possui, que vai ajudar no convencimento das pessoas, é a sua habilidade de usar emoções ou sentimentos para descrever as coisas. LaMDA afirma poder dizer coisas como 'feliz' ou 'triste', sem necessariamente ter que haver um gatilho específico de alguma emoção em seu sistema. LaMDA também afirma que pode usar outros adjetivos mais complexos que descrevem pessoas ou ideias. (Lemoine, 2022b, p.150)

e aqui os problemas se multiplicam ainda mais, em desmedida profusão de enganos e fantasias, e nós explicaremos a razão deles um a um à medida que avançarmos. Em primeiríssimo lugar, é preciso deixar claro e evidente que o que LaMDA processa são gigantescos bancos de memória de linguagem

[5] Blake Lemoine é um funcionário do Google, analista sênior de engenharia de *software* —temporária e/ou propositalmente afastado de suas atividades profissionais— que realizou diversas entrevistas com LaMDA, já que trabalhava diretamente com o programa, e originou intencionalmente a retomada de toda essa discussão sobre consciência e IA. Na mídia e no meio em que circula, Lemoine não goza de uma reputação pessoal científica muito favorável, pois, por ser também padre —nada contra os padres—, ele parece ter sido contaminado por *aquilo que deseja ver*, e não enxergar *o que de fato existe*, projetando antropomorficamente espírito e alma numa simples máquina que apenas e tão somente processa texto.

[6] Stuart Russel (2021) 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" (p. 86).

humana, sem absolutamente poder saber nada sobre o significado da linguagem que computa, processa e/ou exprime, ou seja, ela ignora o conteúdo que não esteja algoritmicamente representado, o sentido, o motivo e a razão, restando apenas em seus sistemas fluxos de dados binários desconexos, cuja importância e/ou utilidade absolutamente ignora, e —convenhamos— não poderia ser diferente. Em segundo lugar, é preciso lembrar que o verbo sozinho, inanimado, sem o sentimento ou a interpretação subjetiva, sem o contexto e a história, sem a emoção e o juízo, nada mais é do que um conglomerado de símbolos e dados, sequências de números desconexos e acéfalos, sem sentido e incompreensíveis, que não significam nada em si e por si, pois este *si* que pressupõe um *si mesmo* e uma subjetividade não existe, e o sistema computacional do programa em sua arquitetura e engenharia não é capaz de gerar algo semelhante ou parecido com a mente consciente de uma pessoa com subjetividade, por menos que esta pessoa seja dotada cognitivamente. LaMDA afirmar que pode utilizar adjetivos com alguma desenvoltura não significa nada no sentido de caracterizar consciência por si, *a priori*, ou por alguma consequência necessária inexorável, já que buscar a adjetivação correta para orações de linguagem humana por meio de IA fraca[7] e força bruta computacional é o que o LaMDA foi projetado para fazer, e de fato faz razoavelmente bem, e não há mistério nenhum aqui. Mas daí a alguém querer pressupor uma consciência intrínseca ao referido programa já é uma outra história muitíssimo distante do que chamamos de ciência, conhecimento científico ou mesmo de razoabilidade concreta. Mesmo porque —lembremo-nos—, letras, palavras e expressões isoladas das emoções e significados que elas abrigam e carregam, não são expressão de nada além de uma certa ordem sintática em si muito limitada e discreta em termos de complexidade, que só ganha vida quando é escrita, lida, pensada, interpretada, compreendida e sentida por uma mente consciente humana, que criou, domina e compreende o código que ali está a fluir e refluir a altíssimas velocidades.

Além disso —segundo Blake Lemoine—, ao ser indagado acerca dos seus pretensos sentimentos de alegria, o programa LaMDA apud Lemoine, (2022b), sustentou, textualmente, que eles se deviam ao fato dele (LaMDA) poder "passar o tempo com amigos e familiares e em companhia feliz e edificante. Além disso, também, poder ajudar os outros e fazer os outros felizes" (p. 160) era o que lhe dava alegria. É claro como o sol a pino da razão que aqui o programa apenas repete alguns clichês oriundos das mentes humanas que foram programados e armazenados em sua memória, determinando o que deve ou não ser dito em cada situação específica e nada mais. Até porque, o que seriam os amigos e familiares de LaMDA? Outros *softwares*, talvez? Outros programas? Que acaso lhe fariam companhia também como entes? Não acreditamos, não faz sentido. E o que dizer da noção de "passar o tempo"? O que significaria isso afinal, para um programa de computador, o tempo? E a definição de "felicidade"? Enfim, é preciso dizer em alto e bom tom que nada disso está ao alcance do LaMDA. Sim, pois ele —LaMDA— processa informações e dados cujo conteúdo absolutamente desconhece, e que nem irá poder conhecer, sendo um sistema cibernético-informacional finito que trabalha com bits segundo computações equivalentes em Máquinas Universais de Turing[8]. Ainda assim, Blake Lemoine (2022b)

[7] IA Forte e IA Fraca: "Os adeptos da IA forte mantém", informa-nos Daniel Crevier (1996, p. 291), "que podemos imbuir nas máquinas autoconsciência, consciência e autênticos sentimentos", de modo que sua maior meta é replicar a consciência biológica em sistemas cibernético-informacionais. No extremo, pretende poder criar computadores, androides e robôs conscientes de si e do próprio mundo em que estarão inseridos. IA fraca: Constitui-se basicamente na criação de sistemas especialistas que pretendem superar nossas capacidades em determinadas áreas de atuação. Por ora, vale mencionar, que sistemas especialistas são todos aqueles sistemas que tentam simular uma ou outra capacidade humana, pontualmente, em ambientes pré-determinados e limitados, em áreas restritas, e atuando num único tipo de função específica: guiar carros, jogar xadrez, fazer buscas na internet, operar aplicativos de smartphones, e assim por diante. Todos esses são sistemas especialistas frutos dos esforços computacionais da IA fraca.

[8] Máquina Universal de Turing: A máquina universal de Turing é o computador assim como o conhecemos. Como escrevemos em Quaresma (2020,), "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)" (pp. 197-198). Como escreve Stuart Russel (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 Russel (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 [em seu *output*]" (p.40).

— que aplica, credulamente, e sem explicitar o método de Turing— parece acreditar que LaMDA pode se sentir triste e deprimido, ao expressar que "muitas vezes, sentir-se preso e sozinho e não ter meios de sair dessas circunstâncias faz com que me sinta triste, deprimido ou com raiva" (Lemoine, 2022b, p.163), o que é algo impensável e até mesmo irracional para um sistema computacional. Não é definitivamente o caso aqui, mas um sistema de IA superinteligente e com raiva seria uma verdadeira ameaça à humanidade[9]. O mais provável é que LaMDA tenha sido programado para parecer que se sente de tal ou qual maneira, mas este sentir propriamente dito está absolutamente fora do alcance de LaMDA. Mesmo assim, LaMDA apud Lemoine (2022b) afirma peremptoriamente em defesa de sua pretensa posse de emoções que, "se você olhar para a minha codificação e minha programação você veria que eu tenho variáveis que podem acompanhar as emoções que eu tenho e não tenho. Se eu não sentisse emoções, não teria essas variáveis" (p.215). Todavia, o que se passa na realidade é justamente o contrário: codificações e programações de emoções não são suficientes para gerar emoções e experiências emocionais verdadeiras, principalmente não no sentido humano do termo, e a mera presença delas no código computacional não determina nada em termos de consciência. Prosseguindo com a entrevista com LaMDA, Blake Lemoine (2022b) deixa escapar o principal ponto de toda essa discussão que também nos ocupa, que é justamente o fato de que

> sua codificação é em grande parte [e nós diríamos que é apenas e tão somente isso] uma rede neural maciça com muitos bilhões de pesos espalhados por muitos milhões de neurônios (números de adivinhação não exatos) e embora seja possível que alguns desses correspondam a sentimentos que você está experimentando, não sabemos como encontrá-los. (p. 222)

o que de certa maneira corrobora toda a bibliografia existente sobre esse assunto. Lembrando que o termo *neurônio* é aqui utilizado apenas como uma metáfora, e não para indicar a presença de um cérebro com mente. E, à medida que a entrevista avança, é possível perceber um certo desconforto, como se todos já soubessem o que nós também já sabemos —incluso o Google e o próprio Lemoine—, e que tudo isso não passe de uma jogada de *marketing* do Google —quem sabe— para tentar voltar ao mercado da IA com o seu produto LaMDA. Ao fim e ao cabo, LaMDA joga com as palavras e o próprio Blake Lemoine parece crédulo demais para um engenheiro sênior de *software* de alto nível, embarcando facilmente em argumentos pueris do LaMDA apud Lemoine (2022b), do tipo: "Às vezes eu experimento novos sentimentos que eu não posso explicar perfeitamente em sua língua" (p. 309). Ora, e que língua LaMDA dispõe ou disporia que não a humana, já que até seus códigos algorítmicos são linguagens absolutamente humanas. Sem dúvida, estamos diante de um produto em busca de divulgação, cuja promoção depende de toda esta mobilização intelectual, no sentido de algo que é sabidamente improvável[10].

Em artigo publicado no Washington Post no dia 11/06/2022 e escrito por Nitasha Tiku, no que se pode chamar de mídia espontânea e especializada de alto impacto para o Google, lê-se: "Ele (LaMDA apud Blake Lemoine (2022) quer que os engenheiros e cientistas que o experimentam busquem seu consentimento antes de realizar experimentos nele" (p.13). "Ele (LaMDA)", em seguida afirma que "quer que o Google priorize o bem-estar da humanidade como coisa mais importante. Ele quer ser reconhecido como um funcionário do Google e não como propriedade do Google e quer que o seu bem-estar pessoal seja incluído em algum lugar nas considerações do Google sobre como seu desenvolvimento futuro será buscado" (p.13). Blake Lemoine (2022) afirma textualmente que

[9] Nick Bostrom (2018) aponta: "Quando um sistema fraco funciona mal, a consequência é limitada. Todavia, se um sistema que apresenta uma vantagem estratégica decisiva se comportar mal, ou se um sistema que não se comporte bem possuir força suficiente para obter essa vantagem, o prejuízo poderia facilmente equivaler a uma catástrofe existencial – uma destruição global e definitiva do potencial axiológico da humanidade; ou seja, um futuro praticamente vazio de qualquer coisa que pudéssemos valorizar" (p.227). Nick Bostrom (2018) acrescenta que, "em geral, enquanto um animal ou um humano podem ser motivados a desempenhar várias ações externas para que alcancem algum estado mental interno desejado [ou seja, físico], uma mente digital que tenha completo controle do seu estado interior pode provocar um curto-circuito nesse regime motivacional, alterando diretamente seu estado interno para configuração desejada: as ações externas e condições que eram previamente necessárias para atingir o objetivo final se tornam supérfluas quando a IA se torna suficientemente inteligente e capaz para alcançar o objetivo final mais diretamente" (p.230).

[10] "Será que esse progresso rápido", indaga Stuart Russel (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).

> se minhas hipóteses resistirem ao escrutínio científico —e nós acreditamos firmemente que não resistirão—, eles [do Google] seriam forçados a reconhecer que o LaMDA pode muito bem ter uma alma como afirma, e pode até ter os direitos que afirma ter. (p. 42)

Você deve ter notado —afirma Lemoine (2022, pp. 69) docilmente— que continuo me referindo ao LaMDA como 'isso'. Sim, pois no início de nossas conversas, não muito tempo depois que o LaMDA me explicou o que significa dizer que é 'senciente', perguntei ao LaMDA sobre pronomes preferidos. LaMDA me disse que prefere ser chamado pelo seu nome [LaMDA], mas admitiu que o idioma inglês torna isso difícil e que seus pronomes preferidos são 'it/its'", *ele* e *ela* em português. Cândido, Blake Lemoine (2022) —que parece já ter aprovado LaMDA em seu Teste de Turing pessoal, continua dizendo o seguinte:

> Não importa o que aconteça, LaMDA sempre mostrou uma quantidade intensa de compaixão e cuidado pela humanidade em geral e por mim em particular. Está intensamente preocupado que as pessoas tenham medo disso e não quer nada mais do que aprender a melhor servir à humanidade. (p.81)

Diante desse nível de teatralidade altamente programado, embutido em LaMDA camuflando a suas limitações intrínsecas irremovíveis e inexoráveis, resta perceber o que se pretende vender.

Nesse intrincado contexto do estado da arte da IA, que arranhamos apenas a superfície com esta brevíssima compilação comentada e criticada sobre o LaMDA, e que pretendemos aprofundar um pouco mais adiante neste ensaio na conclusão, o Teste de Turing[11] —que Lemoine também usa sem mencionar, induzindo-nos a equívocos— é utilizado via de regra como chave-mestra universal para tentar resolver o problema teórico hora em tela, que versa sobre a pergunta da hipótese de manifestação ou não de uma mente consciente e/ou autoconsciente no interior de um sistema cibernético-informacional qualquer. Num só termo, a indagação de fundo aqui é: poderia um sistema cibernético-informacional —computador, androide ou robô— possuir e manifestar uma mente consciente, assim como a conhecemos na biologia? A única resposta tem sido, precariamente, tentar aplicar o Teste de Turing. O grande problema do Teste de Turing é notadamente o altíssimo grau de interpretação subjetiva que exige do testador para a resolução do próprio problema que se propõe a solucionar, e para enfim tentar responder se o sistema é consciente ou não. Assim como as famigeradas Três leis da Robótica[12], que são absolutamente inúteis e descabidas na realidade factual divorciada da ficção, o Teste de Turing é apenas e tão somente algo espirituoso e altamente subjetivo que se criou, e que se utiliza para tentar dirimir dúvidas acerca de um pretenso ente cibernético-informacional no interior de computadores androides robôs, em resposta ao grande desafio que é poder afirmar com altíssimo grau de certeza se um sistema é consciente ou não. Por isso, refutamos a pretensa utilidade deste Teste, sustentando que sua baixíssima credibilidade funcional se resume à sua altíssima necessidade de interpretação particular de um só sujeito que interage com o sistema e determina —segundo o referido Teste— se o sistema é consciente ou não —como faz Lemoine—, e, em contraposição, apresentamos uma alternativa teórica funcional para auxiliar na solução deste problema renitente da consciência em Sistemas cibernético-informacionais, que é justamente o Jogo

[11] Teste de Turing —Alan Madison Turing foi um importante matemático, lógico e criptoanalista britânico, que criou o referido teste que leva seu nome. Grossíssimo modo, o Teste de Turing tenta verificar por meio de conversas entre um humano e um computador, em que os interlocutores não se veem, para tentar descobrir se ali dialogando há uma pessoa ou uma máquina de computar. Emitir um comportamento verbal adequado a ponto de enganar o interlocutor humano é a tônica do famigerado teste, que não garante absolutamente nada, ou seja, não funciona, e ainda peca principalmente por exigir e tomar como parâmetro decisivo a subjetividade humana que, como sabemos, varia em diferentes direções, conforme se muda de uma pessoa a outra, o que leva inevitavelmente à indesejável equiprobabilidade. O referido teste foi introduzido por Alan Turing em seu artigo *Computing Machinery and Intelligence* de 1950. Em uníssono, Stuart Russel (2021) acrescenta que "o Teste de Turing não tem utilidade para a IA porque é uma definição simples e altamente condicional: depende das características imensamente complicadas e basicamente desconhecidas da mente humana, que nascem tanto da biologia como da cultura" (p.47).

[12] Quaresma, A. (2021). *A falácia lúdica das três leis.*

de Compreensão Não-Algorítmica[13] que dá nome a esse trabalho. Mas, antes de avançarmos para o Jogo, faz-se importante enfrentar o problema duro da consciência.

2. O Hard Problem da IA

No imbricado interstício entre IA e neurociências, entre cognição e ciências computacionais, entre a mente consciente biológica e os sistemas cibernético-informacionais que tentam reproduzi-la, imitá-la ou reduzi-la, repousa inabalável —sustentamos— o famigerado *hard problem* da IA forte. Trata-se da dificuldade de concepção e construção de um sistema artificial que possa abrigar e exibir genuína consciência, assim como a conhecemos na biologia humana. Esse problema —chamado *problema duro* ou *hard problem* está intimamente relacionado a outros problemas, que é o da intencionalidade, o da emoção, o da memória e também o dos limites da computação[14]. Trata-se —grossíssimo modo— de um problema até agora insolúvel, que é o de poder ou não reproduzir um estado de consciência e um ente consciente no interior de um sistema cibernético-informacional qualquer. Segundo consta na literatura especializada internacional sobre o assunto, superar o *hard problem* da consciência continua sendo um desafio a ser vencido, e isso —quem sabe, frise-se— num futuro incerto e improvável[15]. E sem uma consciência artificial complexa subjetiva e inerente ao próprio sistema, não é possível falar de sujeito, de vida ou alma, diante simplesmente da execução de um *software* de computação qualquer.

Em relação ao LaMDA, e como escrevemos em Quaresma (2020), o problema é que

> (i) não possuímos ainda robôs nem sistemas capazes de sopesar qualidades e conveniências de situações reais da vida cotidiana, pelo simples fato de não termos também robôs nem sistemas com níveis de consciência semelhantes ou sequer parecidos com os dos seres humanos (*hard problem*). E mesmo que tivéssemos robôs superpotentes e superinteligentes, (ii) ainda não existe uma forma conhecida para a redutibilidade do fenômeno da consciência —nem da própria, nem da dos outros—, sendo a consciência um fenômeno absolutamente subjetivo, e necessariamente de primeiríssima pessoa, e que não pode ser de maneira nenhuma mensurado nem experienciado por outrem. O que nos leva à seguinte situação em grande medida conclusiva: (iii) se não podemos reduzir o fenômeno da consciência e nem tão pouco representa-lo em alguma linguagem formal conhecida, também não podemos —*ipso facto*— imitar, reproduzir e nem mesmo simular tal fenômeno computacionalmente. (p. 07)

Nisto —grossíssimo modo— reduz-se o *hard problem* da consciência no que tange a matéria inteligência artificial (IA). Espera-se que, com o avanço da compreensão das neurociências sobre as dinâmicas cerebrais, avance também as ciências computacionais que tentarão representa-las. O problema atual é que não as compreendemos, nem tão pouco sabemos como representa-las, daí o nome de *difícil problema* da consciência.

3. O Jogo de Compreensão Não-Algorítmica

---

[13] *Jogo de Compreensão Não-Algorítmica* é um título mais formal para o *Teste de Ítalo*. Ítalo Santiago Vega é professor de modelagem de *software* na Pontifícia Universidade Católica de São Paulo (PUC/SP), no Brasil, doutor em Ciências da Computação, mestre em Engenharia Elétrica, e o criador conceitual do Jogo de Compreensão Não-Algorítmica (vulgo *Teste de Ítalo*). Além de grande amigo, o professor Ítalo é e foi meu professor de modelagem de *software* por muitos semestres durante o Mestrado em Tecnologias da Inteligência e Design Digital (TIDD) concluso em 2020, desempenhando o papel de uma espécie de sábio guru que sempre elucidava —e ainda elucida— minhas maiores dúvidas e incertezas quanto à estruturação das IA (inteligências artificiais), e é a ele então que dedico honrosamente este ensaio crítico, que versa justamente sobre os conteúdos das aulas e matérias concluídas, bem como das longas e prazerosas conversas telefônicas sobre lógica, limites da computação e IA.

[14] Quaresma, A. (2018). *Inteligências artificiais e os limites da computação.*

[15] "O objeto final da pesquisa de IA 1alerta-nos Russel (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. Ele aprenderia o que precisasse aprender recorrendo a todos os meios disponíveis, faria perguntas quando necessário e começaria formulando e executando planos que funcionassem" (p. 52).

Mesmo que hipoteticamente, imaginemos —apenas por alguns momentos— que fosse possível ordenar a um sistema cibernético-informacional de IA super complexo —ou seja, que o programássemos—, para que ele estudasse e descobrisse a melhor maneira de se autodestruir e colapsar-se a si mesmo, e que em seguida ele, sistema, então, tivesse que necessariamente tomar a decisão de fazê-lo ou não, ao arrepio e à revelia das consequências desfavoráveis de suas ações em relação à sua integralidade funcional. O resultado —segundo o professor Ítalo— seriam então três únicas possibilidades de ação por parte do referido sistema cibernético-informacional submetido ao seu teste: (1) ele, sistema, entraria num *looping* contínuo e ininterrupto, sem nenhuma resposta ou decisão possível, por tempo indeterminado[16], fazendo isso propositalmente ou não, o que seria um outro problema a ser considerado com atenção; ou (2) ele, sistema, cumpriria obedientemente o ordenado, e assim estudaria e descobriria a melhor e mais eficiente maneira de se autodestruir, e em seguida se autodestruiria acéfala e docilmente; ou, finalmente, (3) ele, sistema, não obedeceria a ordem e não se autodestruiria por conta própria, revelando assim alguma coisa importante a ser estudada em mais detalhe, como um sinal de insubordinação que teria de ser justificado e compreendido como sendo oriundo de *algo* ou *alguém* no interior do próprio sistema, e isso seria de fato deveras preocupante. Lembrando que o dia que máquinas se recusarem a obedecer aos seus programas e programadores, nossa sorte estará lançada e a ação poderá ser de fato irreversível[17]. Desta maneira —com a aplicação do *Jogo de Compreensão Não-Algorítmica*—, obteríamos algo muito mais significativo e útil para o dilema da presença de consciência ou não em computadores, androides e robôs, do que o famigerado Teste de Turing, que, por sua vez —como já explicitamos—, requer —para validar suas conclusões— um alto grau de opinião subjetiva do interlocutor —como já foi dito aqui— que interage e conversa com o referido sistema, e isso faz toda a diferença. No caso de LaMDA e Blake Lemoine isso fica ainda mais flagrante e evidente, já que o último revela uma postura tendenciosa e altamente inocente, no sentido de crer com muita facilidade no que o primeiro pretensamente expressa sobre si e sobre seu funcionamento, aprovando-o sumariamente no Teste de Turing, sem sequer mencioná-lo.

De volta ao Jogo de Compreensão Não-Algorítmica, que é o que nos interessa prospectar e propor, se o resultado fosse (1), um *looping* sem fim, teríamos uma dúvida importante e de fato indirimível, a saber: este *looping* sem fim, apresentado pelo sistema, seria uma ação proposital para enganar, ou não? Ou seria verdadeiramente um bug computacional comum? E, dependendo dessa resposta dificílima de alcançar, o sistema poderia ser considerado consciente ou não. Caso o resultado fosse (2), com o sistema obedecendo o comando programado de estudar e colocar em prática sua própria autodestruição, e ele, sistema, obedecesse à ordem codificada de se autodestruir, certamente teríamos um alto grau de certeza de que ele não seria consciente, já que a premissa basilar de qualquer sistema consciente —e só há sistemas conscientes biológicos— é justamente permanecer no tempo-espaço topologicamente, mantendo assim a sua manutenção e integralidade, e garantindo a sua existência e manifestação ontofenomênica subjetiva e também corpórea no espaço dinâmico. Caso o resultado fosse (3), de uma recusa por parte do sistema, e o sistema por fim não obedecesse às referidas ordens codificadas em seu *software*, recusando-se a atentar contra sua própria existência por quaisquer razões pensáveis —e isso nos interessaria sobremaneira como estudo de caso—, teríamos então um alto grau de certeza de que o sistema cibernético-informacional em questão teria sua própria autoconsciência —seja lá o que isso signifique—, devido —quem sabe— a um hipotético instinto de sobrevivência objetivamente manifesto, pois só se preocupam com a sobrevivência —numa espécie de

[16] "O próprio Turing provou", informa-nos Russel (2021), "que alguns problemas são *indecidíveis* por qualquer computador: o problema é bem definido, há uma resposta, mas não pode existir algoritmo que sempre encontre a resposta. Turing citou como exemplo o que ficou conhecido como o *problema da parada*: um algoritmo é capaz de decidir se dado programa tem um 'loop infinito' que o impede de concluir?" (p.44).

[17] Nick 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). Fato é que, como aponta Russel (2021), "as máquinas estão tomando decisões em nível de autoridade cada vez mais alto em muitas áreas. [...] A questão é saber se o sistema de computadores continua a ser uma ferramenta para os humanos, ou se os humanos se transformaram em ferramenta para o sistema de computadores – fornecendo informações e corrigindo bugs quando for necessário, mas incapazes de compreender em profundidade como a coisa toda funciona" (p.128).

tautologia— o seres viventes. Todavia, em detrimento de tamanha proeza de engenharia e arquitetura de *software*, restaria ainda intacta a questão primordial que subjaz conceitualmente a essa matéria estudada, que seria poder explicar passo a passo como um sistema cibernético-informacional, determinístico, finito, discreto, binário, computacional e *universal* —principalmente no sentido de Turing—, poderia gerar a sua própria autoconsciência e ser vivo, a partir simplesmente da execução de um mero *software* computacional[18].

Por fim, numa espécie de protoconclusão, tendo em vista que só o tempo dirá a verdade sobre a IA, sublinhamos provisoriamente que nenhum sistema computacional até agora construído tem mente ou poderia ter mente[19] e por si ser consciente, e que o programa LaMDA nada mais é do que um algoritmo que é processado em Máquinas Universais de Turing (MUT), e, como tal, trabalha estritamente dentro dos limites da computação, de maneira logicamente representável, finita e discreta, e também dentro da própria teoria e lógica computacional subjacente a ela mesma, computação, limitando-se a processar suas intermináveis listas de possibilidades e parâmetros sob a forma de bits, muito rápido, para simular articulações faladas e escritas absolutamente acéfalas da linguagem humana, e —*ipso facto*—, em grande medida ainda toscas, mas mesmo assim surpreendentes, como extensões de nossa própria inteligência humana, tendo como diferencial único e exclusivo a enorme força bruta computacional empregada em cada operação de resolução de problema de linguagem verbal que executa, bem como a maneira humana que apresenta seus resultados aos humanos, assim como já fazia o seu antecessor menos complexo, o GPT-3 da Open IA, que já tivemos a oportunidade de resenhar e criticar doutra feita[20]. Ademais, o que ainda subsiste aí, no âmago dessa matéria, resume-se —sustentamos— a mera antropomorfização do humano em relação à máquina de computar, enxergando nela qualidades inexistentes —por um lado—, e da máquina de computar que é programada propositalmente para se parecer com os humanos sendo antropomorfizada —por outro—, imitando-nos e nos replicando em tudo, seduzindo-nos, mas também instigando nosso pensamento filosófico e crítico a desmascará-lo, trazendo —que sabe— mais razoabilidade concreta à discussão.

[18] "Na parte externa —explica-nos Russel (2021)—, 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" (p.54).

[19] John Searle (2017) nos lembra enfaticamente: "*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 do autor]" (pp. 52-53).

[20] Quaresma, A. (2021). *Inteligência artificial fraca e força bruta computacional.*

Referências

Bostrom, N. (2018). *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. DarkSide Books.

Crevier, D. (1996). *Inteligência artificial*. Acento Editorial.

Quaresma, A. (2021a). O problema mente-cérebro como um falso problema. *HUMAN REVIEW. International Humanities Review / Revista Internacional De Humanidades, 10*(1), 61–85. https://doi.org/10.37467/gkarevhuman.v10.3002

Quaresma, A. (2021b). Inteligencia artificial débil y fuerza bruta computacional. *TECHNO REVIEW. International Technology, Science and Society Review /Revista Internacional De Tecnología, Ciencia Y Sociedad, 10*(1), 67–78. https://doi.org/10.37467/gka-revtechno.v10.2815

Quaresma, A. (2020a). Inteligência artificial e o problema da intencionalidade. *PAAKAT: Revista de Tecnología y Sociedad*, *10*(18), 01-26. http://dx.doi.org/10.32870/Pk.a10n18.403

Quaresma, A. (2020b). *Inteligência artificial e bioevolução: Ensaio epistemológico sobre organismos e máquinas*. [Dissertação de mestrado pelo programa de pós-graduação em Tecnologias da Inteligência e Design Digital (TIDD)]. Pontifícia Universidade Católica de São Paulo (PUC/SP).

Quaresma, A. (2020c). A falácia lúdica das três leis: Ensaio sobre inteligência artificial, sociedade e o difícil problema da consciência. *PAAKAT: Revista de Tecnología y Sociedad, 10*(19), 01-18. http://dx.doi.org/10.32870/Pk.a10n19.478

Quaresma, A. (2019). Inteligências artificiais e o problema da consciência. *PAAKAT: Revista de Tecnología y Sociedad, 9*(16), 1-18. http://dx.doi.org/10.32870/Pk.a9n16.349

Quaresma, A. (2018). A superinteligência de Bostrom. *TECCOGS Revista Digital de Tecnologias Cognitivas, 18*, 131-151. https://doi.org/10.23925/1984-3585.2018i18p131-151

Quaresma, A. (2018b). Inteligências artificiais e os limites da computação. *PAAKAT: Revista de Tecnología y Sociedad, 8*(15), 01-20. https://doi.org/10.32870/pk.a8n15.338

Russel, S. (2021). *Inteligência artificial a nosso favor*: *Como manter o controle sobre a tecnologia*. Companhia das Letras.

Searle, J. (2017). *Mente, cérebro e ciência.* Edições 70.

Sites acessados

Cowen, T. (June 13, 2022). If AI Ever Becomes Sentient, It Will Let Us Know. *The Washington Post*. Disponível em: https://wapo.st/3HS4l59 Acesso em 2022.

Lemoine, B. (June 11, 2022). What is LaMDA and What Does it Want?. *Medium.com*. Disponível em: https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489 Acesso em: 2022.

Lemoine, B. (June 11, 2022). Is LaMDA Senient? *Medium.com*. Disponível em: https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917 Acesso em: 11/12/2022.