

## O cérebro e o robô: inteligência artificial, biotecnologia e a nova ética

João de Fernandes Teixeira<sup>1</sup>

TEIXEIRA, João de Fernandes. **O cérebro e o robô**: inteligência artificial, biotecnologia e a nova ética. São Paulo: Paulus, 2015, 158p.

Resenha por Eduardo Camargo<sup>2</sup>

### Reflexões sobre o biológico e o artificial

#### Inteligência artificial

A inteligência artificial (IA) é inteligência dos agentes artificiais assim como o campo do seu estudo acadêmico. O termo foi introduzido em 1956 como título da conferência *Dartmouth Summer Research Project on Artificial Intelligence* organizada por John McCarthy. A nova disciplina baseia-se na conjectura de que qualquer aspecto da inteligência humana, da linguagem ao uso de conceitos e abstrações, poderia ser descrito de maneira tão exata que uma máquina deveria ser capaz de reproduzi-lo (McCARTHY, 1955). Se Turing perguntava “Pode uma máquina pensar?”, a IA buscava responder essa questão para qualquer instância da inteligência humana.

A ideia de construir agentes artificiais capazes de imitar comportamentos humanos não é recente. Há inúmeras histórias, ou estórias, de artifícios pensados para este fim, dentre os quais, os autômatos de Descartes e a máquina de jogar xadrez de Kempelen são exemplares clássicos. Independentemente do sucesso na realização de tais dispositivos, o simples fato de se pensar a respeito de máquinas capazes de imitar

---

<sup>1</sup> João de Fernandes Teixeira possui graduação em Filosofia pela Universidade de São Paulo (1977), mestrado em lógica e filosofia da ciência pela Universidade Estadual de Campinas (1982) e doutorado (PhD) em filosofia pela University of Essex (Inglaterra). Fez pós-doutorado nos Estados Unidos em 1998, no Center for Cognitive Studies, na Tufts University, sob a supervisão do Prof. Daniel Dennett. Participou do grupo de ciência cognitiva do Instituto de Estudos Avançados da USP. Foi professor na UNESP (campus-Marília) de 1982 a 1991. Em 1992 ingressou na Universidade Federal de São Carlos, e, atualmente é professor titular nessa universidade.

<sup>2</sup> Eduardo Camargo é Engenheiro Mecânico, Pós-graduado em Marketing, Mestre e Pesquisador em nível de Doutorado em Tecnologias da Inteligência e Design Digital.

o humano traz inúmeras questões filosóficas e éticas, principalmente agora quando os avanços tecnológicos apontam para alguma possibilidade de êxito.

O novo livro de João de Fernandes Teixeira, *O Cérebro e o Robô: Inteligência artificial, biotecnologia e a nova ética* é um ensaio que leva o leitor a revisitar os principais fatores teóricos responsáveis pelo estágio atual de desenvolvimento das tecnologias digitais, apontando as consequências éticas desses avanços, e as principais questões que permeiam o debate filosófico a respeito da coexistência entre humanos e máquinas inteligentes. Para tanto, o autor propõe, na introdução, a inspiradora metáfora da navegação marítima como representação da aventura filosófica. No retorno ao porto de origem, o navegante percebe os porões, que deveriam estar repletos de riquezas, vazios. Segundo T. esta frustração é o que torna a viagem fascinante (p.15). Respostas são importantes e perguntas são fundamentais.

Mas, se há uma viagem marítima em vista, o autor (T.) faz o papel do experiente capitão, conduzindo o leitor-viajante através de temas tão complexos quanto os propostos pelo ensaio: inteligência artificial, biotecnologia, singularidade e novas questões éticas que surgem na combinação desses temas.

O primeiro capítulo, segundo o autor, tem como eixo o resgate da obra de Descartes, colocando-o como realizador do trabalho que dá origem não só à metafísica moderna, mas também à ciência e a tecnologia como entendemos hoje (p.29). Seu ponto de partida é a pergunta clássica “Pode uma máquina pensar?” formulada explicitamente só mais tarde por Turing, mas que encontra precursores no pensamento de Descartes e sua concepção de autômatos. T. afirma que a ideia de que computadores só podem fazer o que foram programados para fazer é ultrapassada e que, alguns robôs e programas de computador são capazes de aprender com seus erros e modificar seus próprios comportamentos. Afirma também, em contraponto, que uma máquina pode ser como um cão que sabe andar por uma casa sem se perder, embora nunca possa imaginá-la. Esta parece ainda uma vantagem humana (p.33-34).

Cabe uma ressalva aqui. Sem dúvida, o aprendizado de máquina (*machine learning*) é uma das tecnologias mais eficientes da IA atual. Baseadas em abordagens probabilísticas ou em análise de dados (*data mining*), as técnicas de aprendizado de máquina são capazes de dotá-las de algum grau de aprendizagem (ABU-MOSTAFA,

2012). Mas seria uma aprendizagem genuína a ponto de se afirmar que os computadores não fazem apenas o que são programados para fazer? Esta é uma questão fundamental no campo da computação emergente, já que não é fácil conciliar a ideia de um determinismo contido na programação com a percepção de fenômenos emergentes como o aprendizado genuíno (EL-HANI e QUEIROZ, 2007). Nossa percepção de que máquinas começam a apresentar capacidades emergentes não seria um equívoco provocado pela complexidade dos programas atuais? Ou seja, o programador determina uma gama de estados primitivos, mas não pode prever todos os estados secundários devido à complexidade de arranjos possíveis, o que não quer dizer que a máquina extrapolou sua programação original.

O autor, então, apresenta uma visão panorâmica da ideia de máquina pensante ao longo da história, destacando a posição dualista de Descartes, para quem, mente e corpo são duas substâncias separadas. Assim, para Descartes, um autômato bem construído poderia fazer tudo o que um ser humano faz, mas nunca teria uma alma (p.35-39). Na filosofia do século XX, contrapõe o autor, muitos filósofos da mente passam a uma crítica vigorosa da posição dualista e, nas últimas décadas, teriam se inclinado em direção à neurociência, cuja tendência predominante é a redução da mente ao cérebro. Na filosofia da mente contemporânea, afirma T., o problema metafísico das relações mente-corpo foi reformulado e passou a ser chamado de *hiato explicativo*.

Em seguir, T. aborda a questão da *causação mental* na qual o significado de algumas palavras pode alterar emoções. O autor afirma que os filósofos materialistas têm a tendência a trivializar o assunto, enquanto os filósofos dualistas tendem a considerar a questão insolúvel (p.43). Neste contexto, T. sublinha as afinidades entre o pensamento de Descartes e de Turing. Os autômatos imaginados por Descartes e Leibniz começam a se tornar realidade com a invenção da máquina de Turing. O autor apresenta também a semelhança entre a noção de tempo em Descartes e a sucessão de estados da máquina de Turing (p. 46-49).

O texto de T. é claro e enriquecedor, sempre apresentando objeções e contrapontos às posições apresentadas. Um aprofundamento nas vertentes não dualistas da inteligência das máquinas tornaria a exposição mais complexa. Tanto os

trabalhos de Leibniz, citado de passagem (p.46), quanto os vieses teóricos baseados na semiótica (máquinas semióticas, por exemplo (NÖTH, 2007) poderiam ter sido considerados.

## **Biotechnologia**

No segundo capítulo, T. apresenta reflexões acerca do mundo tecnológico digital, apontando a IA, a biotecnologia e a nanotecnologia como os resultados mais recentes dessa revolução (p. 52). Sua intenção, aqui é apresentar a aversão metafísica ao mundo tecnológico em obras de filósofos como Heidegger. Chegamos a uma encruzilhada tecnológica? Devemos retroceder diante dos riscos dessas novas tecnologias (p. 30).

Muitas questões referentes à primazia da tecnologia no mundo atual são abordadas pelo autor deste livro de 158 páginas, do aquecimento global à privacidade num mundo conectado, passando pelo industrialismo e a abundância de bens de consumo que não levam, necessariamente, à felicidade. Porém, seu propósito principal é apontar as tecnologias que podem trazer consequências ainda mais impactantes: a nanotecnologia, o melhoramento genético humano, a potencialização química da inteligência e da memória e, finalmente, a inteligência artificial (ciência da computação) que teria tornado possível todas as outras. As grandes preocupações do autor residem no fato de que, até agora, as invenções tecnológicas nos têm sido impostas, e acabamos por aceitá-las passivamente. T. aponta, de maneira acertada e conveniente, que devemos cada vez mais discutir com antecedência a disseminação de algumas tecnologias, intervindo democraticamente nos seus caminhos futuros (p.70).

Singularidade é o tema do terceiro capítulo. O termo é emprestado da física. Ele indica um fenômeno extremo como os buracos negros. Na área da tecnologia, a singularidade corresponde ao momento no qual a inteligência artificial suplantaria a inteligência humana. T. apresenta este viés tecnológico do termo desde seu nascimento na ficção científica de Vernor Vinge até sua adoção por Ray Kurzweil, que o popularizou. O aumento exponencial da capacidade computacional estaria relacionado

com a Lei de Moore, que atesta que a quantidade de transistores que podem caber num *chip* dobra a cada dezoito meses. O que faria o poder dos computadores dobrar no mesmo tempo. Para Kurzweil, a inteligência dos supercomputadores deve igualar a humana por volta de 2045. Então, atingiríamos um limiar a partir do qual a quantidade se torna qualidade e com isso surgiria inteligência e consciência (p.74).

Após uma breve reflexão sobre as consequências da singularidade – como a possibilidade de uma superinteligência ser capaz de produzir uma descrição completa de cada ser humano, levando à imortalidade – T. aponta alguns possíveis fatores impeditivos, que os defensores da singularidade parecem ignorar. Ele pergunta (p.79): “Mas, se a singularidade é uma equiparação com a inteligência humana, será ela dissociável da consciência?” Outra questão fundamental da filosofia da mente ainda não solucionada é a relação entre a mente e o cérebro (p.81). Aqui, também valeria um destaque do autor, dando conta de que a Lei de Moore não é consenso, quando aplicada aos *chips* que temos disponíveis hoje. Alguns cientistas acreditam que, por volta de 2020, os transistores alcançarão o limite da miniaturização, determinando uma estagnação na tecnologia baseada em silício. Se a singularidade vier, deverá ocorrer apoiada em outro substrato tecnológico.

O capítulo termina com a discussão de questões filosóficas e éticas de uma possível superinteligência inclusive a mistura dos humanos com máquinas através do aperfeiçoamento genético e das neuropróteses. T. afirma a necessidade de estarmos à altura da tecnologia que criamos para não nos tornarmos obsoletos e questiona o acesso democrático a todos os aperfeiçoamentos nessas áreas (p.85). A versão digital da clonagem de seres humanos não deveria ser proibida da mesma forma que a clonagem original acabou sendo (p.88)? Neste ponto, devido à relevância dessas questões éticas, um maior aprofundamento filosófico seria bem-vindo. Tal fato, no entanto, não diminui a qualidade da exposição panorâmica do autor.

Outra questão que pode ser impeditiva à singularidade diz respeito ao problema do significado, que o autor explora no quarto capítulo sobre o teste de Turing (1950) e o argumento do quarto chinês de Searle (1984). T. inicia o capítulo revisitando dois dos mais representativos e conhecidos experimentos mentais da IA

para questionar a possibilidade de uma máquina ser capaz de pensar e de ser consciente desse pensamento.

Searle conclui, no seu argumento do quarto chinês, que dizer que uma máquina compreende algo é um equívoco, pois os pensamentos ou “estados mentais” de uma máquina são sobre o nada. Elas não têm qualquer relação com o mundo externo. Não podemos pensar a não ser que pensemos em alguma coisa e, a esta capacidade de pensar em algo, Searle chama de intencionalidade (p.95). Sua conclusão é que os estados intencionais são uma representação do mundo e não podem existir de maneira inconsciente. Assim a consciência seria parte fundamental do significado (p.96).

T. afirma que o argumento do quarto chinês se transformou num ícone da crítica à IA e, conseqüentemente à singularidade. A base da intencionalidade e da consciência é a vida. Sem um cérebro vivo, elas não seriam possíveis (p.97). T. argumenta, então que esta posição de Searle pode ser questionada. Afinal, o que é vida? Após reflexões acerca desta questão, conclui em acordo com Daniel Dennett que, devido ao caráter binário do DNA, a vida é um fenômeno computável, o que permitiria construir uma réplica do cérebro humano. Porém, se assim fosse, porque ele não teria estados intencionais? O autor lembra que Searle não descarta essa possibilidade, apenas entende que estamos ainda muito distantes de replicar a matéria viva (p. 100).

No final do capítulo, o autor apresenta a teoria do significado apresentada por Wittgenstein nas *Investigações filosóficas*, que se opõe a Searle. Nesta concepção, o significado é gerado pelas práticas de uso da linguagem e não por estados mentais intencionais associados às palavras e crenças (p.104). A conclusão a partir de então é que o ocupante do quarto chinês não poderia gerar símbolos com significado devido ao caráter privado de sua linguagem. Assim, de uma perspectiva *wittgensteiniana*, o argumento do quarto chinês é redundante, deixando de ser uma fortaleza inabalável. Mesmo assim, T. acredita que dificilmente um opositor da IA aceitaria a crítica baseada em Wittgenstein (p.109).

O desenvolvimento da série de passos argumentativos que levam ao contraponto entre Searle e Wittgenstein é um clímax do livro. De maneira clara e elegante, T. introduz o leitor a ideias chave de Wittgenstein com relevância para o tema do livro.

## Implicações éticas

Os dois últimos capítulos tratam de questões éticas envolvendo o transumanismo, decorrente da associação homem-máquina e da reforma biotecnológica (cap. 5), assim como da IA biológica (cap. 6).

T. apresenta uma perspectiva bastante abrangente acerca da problemática que as novas tecnologias podem trazer ao debate filosófico. Esta parte do texto é provocativa e questionadora. De andróides, como seres orgânicos modificados, aos robôs, como seres totalmente inorgânicos, e passando pelos ciborgues como o meio termo entre os dois, praticamente todas as possibilidades são enfrentadas. Neste percurso o autor nos põe em contato com as leis da robótica discutidas por Isaac Asimov, e com o novo termo “roboética”, criado por Gianmarco Veruggio para designar a ética específica que deverá reger as relações entre humanos e máquinas inteligentes (p.114).

Quando trata do transumanismo como um programa de aperfeiçoamento do ser humano, o autor apresenta as contribuições dos principais cientistas e pensadores envolvidos com o tema, o biólogo Julian Huxley, o geneticista John Haldane e os filósofos Max More, Nick Bostrom e David Pearce (p. 120). Com relação ao conteúdo, T. aponta a nanotecnologia, a engenharia genética e a IA como as tecnologias que, combinadas, darão suporte ao aperfeiçoamento humano. Mas isto trará consequências práticas e éticas. Não só a inteligência será aperfeiçoada, mas também a longevidade. Quais os problemas sociais decorrentes de vivermos muito mais tempo do que hoje? Os avanços tecnológicos estarão disponíveis para todos ou apenas para alguns grupos privilegiados? Quais os riscos com a diminuição da biodiversidade e forte padronização da espécie humana? Estas e outras questões prementes levam o leitor, de forma natural, à reflexão.

Na sua parte final o autor aborda o tema da IA biológica. Trata-se da mistura de humanos com robôs a partir da cultura de neurônios *in vitro*. A proposta da IA biológica não é replicar a natureza, mas se apropriar dela e dar continuidade à evolução através de uma combinação entre seres vivos e próteses (p. 138).

T. apresenta, então, os trabalhos de Kevin Warwick relativos ao cultivo de neurônios de fetos de ratos e sua conexão, através de microelétrodos, a um robô (*miabots*) que oferece informações sobre o seu entorno (p.138). No futuro, a perspectiva é de utilizar essa técnica com neurônios humanos. Apesar do estágio inicial, algumas consequências podem ser imaginadas. Um robô com cérebro idêntico ao humano acoplado a um corpo artificial não desafiaria definitivamente a posição de Searle? Uma máquina como esta não preservaria os poderes causais do cérebro a ponto de gerar estados intencionais e produzir significado? Estas são algumas das questões levantadas por T. neste capítulo, que flerta com a ficção científica. (No filme de 1995, *La cité des enfant perdus* de Marc Caro e Jean-Pierre Jeunet, por exemplo, um cérebro vive num aquário.) Mas, como ele mesmo diz, na introdução ao seu ensaio, “os filósofos profissionais perderam espaço para a ficção científica, que passou a ser o grande guia da agenda futura da ciência” (p.19).

T. lembra que a possibilidade de cultivar um cérebro de proveta ainda é hipotética e que esse tipo de máquina seria, para Heidegger, o humano inumano sem nascimento nem morte e produziria uma enorme quantidade de problemas éticos. As questões aqui tomam a forma de: será que podemos atribuir cidadania política e direitos civis a um cérebro de proveta? O cérebro de proveta poderá ser considerado uma pessoa? O capítulo é encerrado com a oportuna discussão a respeito da bioética e seus dilemas (p. 147).

*O cérebro e o Robô: inteligência artificial, biotecnologia e a nova ética* é um ensaio que percorre as principais questões relativas ao impacto dos avanços da IA e da biotecnologia. O texto é claro e preciso, e expõe as principais consequências éticas e filosóficas que enfrentaremos no futuro, mas que já se impõe de maneira premente. Apesar do caráter panorâmico, a obra é complexa como o tema exige. Desta forma, é uma obra necessária e bem-vinda.

O leitor-viajante, que embarcou na nau proposta pelo autor, chega seguro ao porto de origem, sem esconder o desejo por uma nova aventura filosófica. Méritos ao capitão.



## Referências

ABU-MOSTAFA, Yaser S. et al., **Learning from data: a short course**. [S.l.], AML Book. 2012.

EL-HANI, Charbel Niño e QUEIROZ, João. Estruturalismo hierárquico, semiose e emergência. Em: **Computação, cognição, semiose**, org. QUEIROZ, João et al. Salvador, EDUFBA, pp. 93-128, 2007.

MCCARTHY, John et al. **A proposal for the Dartmouth summer research project on artificial intelligence**. AI Magazine, 27(4):12-14, 1955.

NÖTH, Winfried. Máquinas Semióticas. Em: **Computação, cognição, semiose**, org. QUEIROZ, João et al. Salvador, EDUFBA, pp. 159-183, 2007.

SEARLE, John. **Minds, brains and science: The 1984 Reith Lectures**. St. Ives, Clays, 1984.

TURING, Alan. **Computing machinery and intelligence**. Mind 49:433-460, 1950.