O que é inteligência artificial (IA)?
Análise em três atos de um conceito em desenvolvimento

João Victor Archegas
Gabriella Maia

Resumo

Em geral, quando falamos em inteligência artificial (IA), pensamos automaticamente em tramas de ficção científica, como "2001: Uma Odisseia no Espaço" e "Eu, Robô", que, entretanto, não correspondem à realidade dessa nova tecnologia. Para complicar ainda mais o debate, quando avaliamos a literatura especializada sobre o tema, são poucos os autores que se preocupam com a definição de um conceito. Neste capítulo, então, nosso objetivo é identificar os elementos que podem nos auxiliar a responder uma pergunta elementar: afinal, o que é IA? Em primeiro lugar, oferecemos uma primeira aproximação conceitual e, paralelamente, uma breve retomada histórica do desenvolvimento da IA desde os anos 50. Em segundo lugar, enfrentamos alguns subconceitos do campo de estudo da IA, como *machine learning* e *deep learning*. Por fim, em terceiro lugar, apresentamos como se dá o debate a respeito da IA no plano internacional, destacando alguns princípios que devem guiar a regulação dessa tecnologia.

Abstract

In general, when we talk about artificial intelligence (AI), we automatically think of science fiction plots, such as "2001: A Space Odyssey" and "I, Robot", which, however, do not correspond to the reality of this new technology. To further complicate the debate, when we evaluate the specialized literature on the subject, there are few authors who are concerned with defining a concept. In this chapter, then, our goal is to identify the elements that can help us answer an elementary question: after all, what is AI? First, we offer a first conceptual approach and, in parallel, a brief historical review of the development of AI since the 1950s. Second, we tackle some subconcepts of the AI field of study, such as machine learning and deep learning. Finally, in the third place, we present the debate about AI at the international level, highlighting some principles that should guide the regulation of this technology.

Introdução

Quando falamos em inteligência artificial (IA) nossos pensamentos são rapidamente transportados para os livros e filmes de ficção científica. Um exemplo é o computador HAL 9000, que, com sua lente preta e foco avermelhado, é responsável por controlar todas as funções da nave *Discovery One* em "2001: Uma Odisseia no Espaço", filme dirigido por Stanley Kubrick em 1968. HAL é apresentado ao público como uma sofisticada aplicação de IA que escapa ao controle de seus programadores e, assim, se torna o grande vilão da trama, colocando em risco a tripulação da espaçonave.

O que poucos mencionam, entretanto, é que, no segundo livro da série de Arthur C. Clarke, HAL é reprogramado por seu criador, Dr. Chandra, que prontamente explica aos demais personagens que os eventos de 2001 só ocorreram porque o computador não foi informado do objetivo de sua missão, o que fez com que a IA desenvolvesse uma espécie de paranoia. Ou seja, é restabelecida, na própria narrativa, a su-

premacia humana em relação à máquina. Nada obstante, a cultura pop se desenvolveu em torno da imagem da IA que pode, a qualquer momento, fugir de controle e ameaçar a existência da humanidade.

O potencial destrutivo da IA ou, em outras palavras, o risco existencial representado por essa nova tecnologia é responsável por pautar boa parte da literatura especializada no tema, em especial no campo do Direito e das ciências sociais. Embora parte dessa discussão seja indevidamente distorcida por preocupações exageradas, são cada vez mais importantes as perspectivas que objetivam compatibilizar o desenvolvimento e a implementação da IA com valores e princípios humanos. Ou seja, a IA deve beneficiar e não ameaçar a humanidade em suas diversas dimensões, desde relações de trabalho até o exercício da democracia.

Uma versão dessa discussão também tem sua origem na ficção científica. Em 1942, o escritor Isaac Asimov apresentou as três leis da robótica em um dos contos da sua coletânea "Eu, Robô" (*I, Robot*) com o objetivo de estabelecer uma relação pacífica entre máquinas inteligentes e a humanidade. São elas: (1) Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal; (2) Um robô deve obedecer às ordens dadas por seres humanos exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei; e (3) Um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou a Segunda Lei.[1]

Ainda que as leis sugeridas por Asimov tenham sido influentes para além das páginas de suas histórias, elas são dirigidas aos próprios robôs, e não às pessoas que desenvolvem e implementam novas tecnologias. Buscando preencher essa lacuna, Frank Pasquale sugeriu, em 2020, quatro novas leis da robótica: (1) Sistemas robóticos e a IA devem complementar trabalhadores e profissionais e não substituí-los; (2) Sistemas robóticos e a IA não devem falsificar e imitar a humanidade; (3) Sistemas

1 FARINACCIO, Rafael. Como funcionam as Três Leis da Robótica do escritor Isaac Asimov em 2017? **Tecmundo**, 11 de dezembro de 2017. Disponível em: https://bit.ly/38Z7sIL.

robóticos e a IA não devem intensificar corridas armamentistas de soma zero; e (4) Sistemas robóticos e a IA devem sempre indicar a identidade de seus criadores, controladores e donos.[2]

Embora o debate sobre os riscos e benefícios da IA para a humanidade seja cada vez mais necessário, poucos autores que exploram a temática se preocupam em definir o conceito com o qual trabalham. Afinal, o que exatamente é inteligência artificial? O argumento central do presente ensaio é que o conceito de IA está em constante evolução e, por isso, deve ser definido de uma forma aberta e dinâmica. Isso não significa, entretanto, que não seja possível estabelecer algumas balizas capazes de separar aquilo que o campo de pesquisa e desenvolvimento em IA engloba e aquilo que, pelo menos neste momento, ainda está fora do seu alcance.

Importa destacar, desde já, que as palavras "artificial" e "inteligente" podem causar alguma confusão a respeito da natureza da IA. Como argumenta Kate Crawford, é necessário desconstruir o mito de que estamos falando de um campo meramente técnico; pelo contrário, "a IA é fundamentalmente política"[3]. Nesse sentido, Crawford demonstra que a IA, paradoxalmente, não é artificial – já que é constituída por recursos naturais e humanos – nem inteligente – uma vez que depende de treinamento computacional intensivo, sem o qual não consegue agir de forma autônoma ou racional.[4] Para ela, devemos entender essa tecnologia como uma espécie de atlas, ou seja, uma representação específica do mundo que está sujeita a influências políticas, econômicas, sociais e tantas outras.[5]

Para melhor compreender esse conceito ainda em desenvolvimento, este artigo se divide em três partes. A primeira apresenta uma primeira

2 PASQUALE, Frank. **New laws of robotics**: defending human expertise in the age of AI. Cambridge: Harvard University Press, 2020, pp. 1-19.

3 CRAWFORD, Kate. **Atlas of AI**: power, politics, and the planetary costs of artificial intelligence. New Haven: Yale University Press, 2021, p. 9.

4 *Ibidem*, pp. 7-9.

5 *Ibidem*, pp. 9-14.

aproximação conceitual e, paralelamente, uma breve retomada histórica[6] da IA enquanto tecnologia e campo de investigações científicas, transitando desde a introdução do teste de Turing em 1950 e a conferência na Universidade Dartmouth em 1955 até o estado da arte hoje. A segunda se debruça sobre os diversos subconceitos que estão albergados por essa tecnologia, como, por exemplo, *machine learning*, *deep learning* e redes neurais. Por fim, a terceira apresenta, de forma sucinta, o espectro da IA no plano internacional, destacando os princípios que devem guiar o desenvolvimento e a aplicação da tecnologia.

1. Primeira aproximação conceitual e breve história da inteligência artificial

Em linhas gerais, a IA é um braço da computação cujo objetivo primordial é desenvolver programas computacionais capazes de automatizar ações inteligentes. Naturalmente, pesquisadores que trabalham nessa linha têm como referência a inteligência humana e, assim, buscam desenvolver mecanismos capazes de observar, aprender, pensar e agir como humanos. Tome-se como exemplo o carro semiautônomo da Tesla. Tal como um piloto humano, o veículo tem capacidade para analisar as condições da via na qual se encontra e, após processar esses dados, decidir qual é a velocidade ideal naquele contexto, se é preciso desviar de obstáculos ou então se é o momento certo para ultrapassar outro veículo.[7]

De forma semelhante, a IBM, uma das primeiras empresas a desenvolver e comercializar aplicações de IA em escala industrial, afirma que "em sua forma mais simples a inteligência artificial é um campo que

6 Como qualquer história, a que vamos contar é inevitavelmente incompleta e gira em torno de poucas figuras (quase todos homens brancos do Norte Global) e acontecimentos que, em retrospecto, marcaram o campo da inteligência artificial desde a metade do século passado.

7 THOMPSON, Cadie. Here's How Tesla's Autopilot Works. **Business Insider**, 1 de julho de 2016. Disponível em: https://bit.ly/3LVKf96.

combina ciência da computação com bases de dados robustas para permitir a solução de problemas"[8]. Outras empresas, entretanto, preferem definir IA com base nos produtos que são desenvolvidos, não se referindo apenas ao campo acadêmico no qual esses debates se inserem. É o caso da Oracle ao afirmar que "inteligência artificial se refere a sistemas ou máquinas que imitam a inteligência humana para performar tarefas e que podem, de forma interativa, se desenvolver autonomamente com base nas informações que coletam"[9].

Reunindo diferentes perspectivas em um conceito com mais nuances, o Parlamento Europeu define IA em três etapas.[10] Primeiro, "é a habilidade de uma máquina apresentar capacidades humanas como raciocínio, aprendizado, planejamento e criatividade". Segundo, uma aplicação de IA é capaz de observar o contexto no qual se insere e agir, dentro de suas limitações, para atingir objetivos pré-definidos. Por fim, em terceiro lugar, uma aplicação de IA também pode, de forma autônoma, adaptar suas futuras ações levando em consideração sua experiência passada. Veja-se, entretanto, que ao detalhar o conceito de IA o Parlamento Europeu corre o risco de alienar algumas tecnologias que hoje são compreendidas como parte desse braço da computação mas não atingiram um grau de sofisticação a ponto de, por exemplo, "observarem o seu contexto".

1.1 *"Can machines think?"*

Embora as definições esboçadas acima já apontem para um possível caminho conceitual, é indispensável retomar a história de origem des-

8 IBM Cloud Education. What is Artificial Intelligence? **IBM**, 3 de junho de 2020. Disponível em: https://ibm.co/3Flos8t.

9 Oracle Cloud Infrastructure. What is AI? Learn about artificial intelligence. **Oracle**, 2022. Disponível em: https://bit.ly/3P2uCyu.

10 European Parliament News. What is artificial intelligence and how is it used? **Parlamento Europeu**, 4 de setembro de 2020. Disponível em: https://bit.ly/3ykytRU.

sa tecnologia para compreender todas as suas complexidades. Alguns autores, como é o caso de Nils Nilsson, de Stanford, sugerem que os primeiros passos foram dados por Aristóteles que buscou "codificar" o processo de raciocínio humano e a estrutura dos argumentos com base em "silogismos".[11] Estudantes da lógica aristotélica, séculos depois, tentaram automatizar a inteligência humana. É o caso Ramon Llull, que propôs no século XIII a criação da *Ars Magna*, uma engrenagem capaz de indicar a resposta para todos os problemas.[12]

Ainda que alguns paralelos históricos como esses possam ser traçados, fato é que a ideia de IA como conhecemos hoje só começou a ser desenvolvida entre as décadas de 40 e 50 do século passado. Antes mesmo do termo ser criado, Alan Turing publicou o seu influente artigo *Computing Machinery and Intelligence*, no qual questionava se máquinas poderiam pensar (*"can machines think?"*).[13] Para Turing, uma máquina só poderia ser considerada inteligente se ela conseguisse convencer um interrogador humano de que ela também é humana.[14] Esse exercício ficou conhecido como o teste de Turing e, por muitos anos, pautou os debates a respeito do desenvolvimento de aplicações de IA.

Afinal, para passar no teste uma máquina precisaria de quatro habilidades fundamentais: (1) Processamento de linguagem natural; (2) Representação de conhecimento; (3) Raciocínio lógico automatizado; e (4) Aprendizado de máquina (*machine learning*). Ademais, alguns au-

11 NILSSON, Nils J. **Artificial Intelligence**: a new synthesis. São Francisco: Morgan Kaufmann Publishers, 1998, p. 8.

12 CARLES SIERRA, Alexander Fidora (Ed.). **Ramon Llull**: From the Ars Magna to Artificial Intelligence. Barcelona: Artificial Intelligence Research Institute, 2011.

13 TURING, Alan. Computing Machinery and Intelligence. **Mind** 49, 433-460, 1950.

14 Essa é uma versão simplificada do teste que Turing propõe em seu artigo, o qual envolve mais atores e etapas. Hoje vários sistemas de IA já são capazes de enganar interlocutores humanos, que acreditam estar interagindo com outras pessoas e não máquinas. Nesse sentido, Nils Nilsson afirma que o teste simplificado perdeu seu valor. Para uma discussão detalhada do teste original, ver NILSSON, Nils J. **Artificial Intelligence**: A new synthesis. São Francisco: Morgan Kaufmann Publishers, 1998, pp. 6-7.

tores propuseram uma versão absoluta do teste de Turing, para o qual a máquina precisaria interagir com pessoas e objetos no mundo físico. Outras duas habilidades seriam necessárias nesse cenário: (5) Visão de computador e (6) Robótica. Como afirmam Stuart Russel e Peter Norvig, essas são as "seis disciplinas que compõem a maior parte da IA".[15]

1.2 Conferência de Dartmouth

Enquanto Turing foi o primeiro a formular um teste para averiguar se uma máquina é ou não inteligente, promovendo, assim, uma verdadeira revolução na computação, John McCarthy, na época em Dartmouth College, é comumente creditado pela criação do termo *artificial intelligence*, que usou como o título de uma conferência organizada em seu departamento no ano de 1956. É curioso notar que diversos outros nomes foram considerados por McCarthy e seus colegas, como *complex information processing*, *machine intelligence*, *heuristic programming* e *cognology*.[16] Nenhuma das alternativas, entretanto, teve a mesma aderência que inteligência artificial.

Ou seja, o termo que usamos hoje para se referir a esse braço específico da computação é, em parte, resultado de um fenômeno conhecido como *path dependence* ou dependência de trajetória. Para Russel e Norvig, o termo *computational rationality* seria mais preciso, mas McCarthy, por uma questão de diplomacia, evitou qualquer menção a computadores no título da conferência porque Norbert Wiener, um dos participantes, focava no uso de dispositivos cibernéticos analógicos ao invés de computadores digitais.[17] Não fosse a pesquisa de Wiener, hoje o título deste ensaio poderia ser "O que é Racionalidade Computacional (RC)?".

15 RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. Pearson, 2020, p. 2.

16 NILSSON, Nils J. **Artificial Intelligence**: a new synthesis. São Francisco: Morgan Kaufmann Publishers, 1998, p. 8.

17 RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. Pearson, 2020, p. 18.

Anos depois, em 2004, McCarthy publicou um artigo para responder algumas questões básicas sobre o tema. Para ele, IA "é a ciência e engenharia de construir máquinas inteligentes, especialmente programas de computador inteligentes".[18] Essa definição pode ser encarada como um indício que, com o passar dos anos, McCarthy foi convencido da centralidade dos computadores digitais para o desenvolvimento da IA. Mesmo diante dessa questão conceitual, a proposta elaborada por McCarthy e enviada à administração de Dartmouth na década de 50 ajuda a compreender a moldura que os pesquisadores buscavam dar ao campo:

> O estudo vai proceder com base na conjectura que todos os aspectos de aprendizagem e quaisquer outros elementos da inteligência podem, em princípio, ser precisamente descritos de forma que uma máquina seja capaz de simulá-los. Será feita uma tentativa de descobrir como fazer máquinas usarem linguagem, formar abstrações e conceitos, resolver problemas que hoje são reservados a humanos e melhorar a si mesmas (tradução livre).[19]

1.3 Que os jogos comecem

Uma significativa parte das descobertas envolvendo IA de 1950 para cá ocorreu durante o desenvolvimento e implementação de programas pensados para jogos, em especial o xadrez. A lógica pode ser resumida da seguinte forma: uma partida de xadrez envolve uma série de habilidades humanas que podem ser classificadas como inteligentes e, por isso, é um objeto de estudo adequado para melhor compreender nossos mecanismos intelectuais e, portanto, permitir o desenvolvimento de outras aplicações mais complexas de IA. Em outras palavras, como

18 MCCARTHY, John. What Is Artificial Intelligence? 2004. Disponível em: https://bit.ly/3ykUyPW.

19 RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. Pearson, 2020, p. 18.

falou uma vez o pesquisador russo Alexander Kronrod, "o xadrez é a *Drosophila* da IA".[20]

Esperançoso com as possibilidades que a IA apresentou à comunidade científica após a conferência de Dartmouth, Herbert Simon fez, em 1957, a previsão que dentro de uma década um programa de computador seria capaz de derrotar o campeão mundial de xadrez.[21] Embora Simon tenha sido otimista demais, sua visão se concretizou com algumas décadas de atraso em 11 de maio de 1997, quando o programa *Deep Blue* da IBM venceu uma série de seis partidas contra o então campeão mundial Garry Kasparov. O enxadrista disse recentemente em entrevista que, embora tenha sido uma experiência desagradável, o episódio o ajudou a compreender o futuro da colaboração entre máquina e humanidade.[22]

Entretanto, o xadrez era apenas um ponto na trajetória da IA, não o seu destino. Como McCarthy já havia antecipado em 2004, a próxima fronteira seria vencer um jogador profissional de Go, um jogo de tabuleiro popular na Coreia do Sul, na China e no Japão. Nas suas palavras, "[o] Go expõe as fragilidades do nosso atual conhecimento dos mecanismos intelectuais envolvidos em jogos".[23] Isso se deve ao fato de que o tabuleiro de Go é maior do que o de xadrez e muitas jogadas são classificadas como instintivas em vez de puramente racionais e calculadas. Ainda assim, em 2016 a IA *AlphaGo* desenvolvida pela empresa DeepMind venceu o campeão Lee Se-dol, que se aposentou em 2019 e disse: "Mesmo que eu me torne o número um, existe uma entidade que não pode ser vencida".[24]

20 MCCARTHY, John. What Is Artificial Intelligence? 2004, p. 6.

21 RUSSELL, Stuart. NORVIG, Peter. **Artificial Intelligence: A modern approach**. Pearson, 2020, p. 21.

22 KNIGHT, Will. Defeated chess champion Garry Kasparov has made peace with AI. **WIRED**, 21 de fevereiro de 2020. Disponível em: https://bit.ly/3PeR08a.

23 MCCARTHY, John. What Is Artificial Intelligence? 2004, p. 6.

24 VINCENT, James. Former Go champion beaten by DeepMind retires after declaring AI invincible. **The Verge**, 27 de novembro de 2019. Disponível em: https://bit.ly/3MYu1fr.

1.4 Universo em expansão

Hoje já existem inúmeras aplicações de IA que vão muito além do xadrez ou Go. O desenvolvimento de novos sistemas de IA foca, basicamente, em cinco dimensões ou habilidades humanas que são comumente associadas à inteligência. São elas: (1) Aprendizagem (automatizada por erro e acerto ou, de forma mais sofisticada, por generalização); (2) Raciocínio (que pode ser subdividido em sistemas dedutivos e indutivos); (3) Solução de problemas (quando um sistema avalia um universo de ações e adota aquela que conduz à concretização de um objetivo pré-definido, como a vitória em uma partida de xadrez); (4) Percepção (veja-se o exemplo trabalhado anteriormente do carro semiautônomo da Tesla); e (5) Domínio da linguagem (é o caso dos processadores de linguagem natural, como o avançado GPT-3 da OpenAI, e os mais comuns *chatbots*).[25]

2. Uma constelação de concepções

Ainda que o conceito de IA esteja em constante evolução e careça de uma definição consensual tanto entre os doutrinadores como os *players* do mercado, podemos observar alguns subconceitos melhor delimitados que integram a constelação de concepções que compõem a inteligência artificial enquanto campo de investigação. Entender esses elementos que compõem a IA é fundamental para melhor compreender essa nova tecnologia em sua amplitude, na busca por uma definição mais adequada e completa.

25 COPELAND, B.J. Artificial Intelligence. **Britannica**, 18 de março de 2022. Disponível em: https://bit.ly/39M9Cfs.

2.1 Subcampos da IA: *Machine Learning*, *Deep Learning* e redes neurais

Existem, em geral, dois subcampos da IA: *machine learning* (aprendizado de máquina) e *deep (machine) learning* (aprendizagem profunda).

O *machine learning* foca no uso de dados e algoritmos para imitar a forma com que humanos aprendem, gradualmente melhorando sua precisão ou acurácia.[26] Na prática, é o campo responsável pelo treinamento de algoritmos para que possam, com eficiência, responder perguntas e resolver problemas a partir do processamento de um grande volume de dados, a exemplo da implementação de árvores de decisão.

Já o *deep learning* é, na prática, um subcampo mais sofisticado e avançado do *machine learning*. Para Lex Fridman, do MIT, *deep learning* seria um "*machine learning* escalável"[27], a partir da qual a IA aprende a executar tarefas mais complexas, utilizando, em regra, redes neurais (*neural networks*).

Redes neurais artificiais (ANNs na sigla em inglês), por sua vez, foram um subcampo do próprio *deep learning*. Correspondem a uma estrutura sistemática interconectada (motivo de sua nomenclatura, uma referência às conexões neurais múltiplas das sinapses), com nós (ou neurônios artificiais) que correspondem a dados conectados uns aos outros em diversas camadas sobrepostas. Daí a ideia de "profundo" ou *deep*.[28] A interconexão múltipla desses dados resulta num aprendizado para execução de tarefas mais complexas.

26 IBM Cloud Learn Hub. What is Machine Learning? **IBM**, 15 de julho de 2020. Disponível em: https://ibm.co/396bwqG.

27 "Scalable machine learning", conforme abordado em palestra no MIT, disponível em: https://www.youtube.com/watch?v=O5xeyoRL95U.

28 IBM Cloud Learn Hub. What is Machine Learning? **IBM**, 15 de julho de 2020. Disponível em: https://ibm.co/396bwqG.

2.2 Aprendizado: supervisionado, não supervisionado e por reforço

As principais técnicas utilizadas para o desenvolvimento de IA se dividem em (1) aprendizado supervisionado, (2) aprendizado não supervisionado e (3) aprendizado por reforço.

O aprendizado supervisionado, técnica mais utilizada hoje, se dá a partir de um grande volume de dados previamente rotulados por uma espécie de etiqueta que identifica o que cada dado representa. Acontece, por exemplo, quando um conjunto de dados sobre transações de cartão de crédito possui um campo específico que indica se ocorreu ou não fraude. A partir dessa rotulação é possível, com o tempo, treinar um sistema antifraude para classificar compras futuras.[29]

Outra técnica que também depende de um grande volume de dados é o aprendizado não supervisionado, que se dá a partir de dados não rotulados. É utilizada, por exemplo, para a identificação de perfis de consumo. Através da coleta de um grande volume de dados de consumidores de um determinado setor da economia, é possível, pelo processamento automatizado por um algoritmo, descobrir padrões de consumo e usar esse *output* para tomar decisões estratégicas de mercado.

Ainda, há o aprendizado por reforço, que se dá através de erros e acertos a partir da aplicação de testes A/B. É aplicado, por exemplo, nas publicidades direcionadas aos usuários em redes sociais. Conforme a interação com um novo conteúdo, em comparação a outros com os quais o usuário também interage, se dá um teste A/B: conteúdos semelhantes sobre temas que interessaram o usuário no passado serão cada vez mais exibidos no futuro. É o que acontece também em serviços de *streaming*,

29 CORTIZ, Diogo. Inteligência Artificial: equidade, justiça e consequências. **Panorama Setorial da Internet**, ano 12, n. 1, 2020, p. 2. Disponível em: https://bit.ly/399361K.

que apresentam indicações de filmes e séries conforme os padrões de consumo de cada usuário.[30]

2.3 Maturidade e níveis dos sistemas: IA Forte, IA Fraca e Superinteligência Artificial

Outra classificação possível para IA, ainda que incompleta, seria quanto à maturidade do sistema. Essa classificação divide-se em três níveis: (1) IA fraca ou restrita (*Artificial Narrow Intelligence* ou ANI), (2) IA forte (*Artificial General Intelligence* ou AGI) e (3) Superinteligência Artificial (*Artificial Super Intelligence* ou ASI).

Uma IA fraca seria um sistema programado para lidar com tarefas singulares ou limitadas[31], podendo até executar tarefas complexas, mas sempre voltadas ao objetivo para o qual foram programadas. É o caso de todas as aplicações comerciais de IA que conhecemos hoje, as quais simulam a inteligência humana mas não possuem autoconsciência, como assistentes de voz em nossos *smartphones*, o computador IBM Watson[32] e até mesmo veículos autônomos.

Uma IA forte, por sua vez, é capaz de lidar com uma ampla quantidade de tarefas concomitantemente. É um sistema que, ao menos teoricamente, pensa e age tal qual um humano graças ao aprendizado por técnicas de *machine learning* e *deep learning*. Essa IA poderá ser: (a) uma máquina ciente, que compreende os estímulos para processar informações ou (b) uma máquina autoconsciente, ou seja, consciente do mundo e de si mesma. É importante pontuar que, hoje, a IA forte é um conceito predominantemente teórico.

30 Para entender melhor o funcionamento dos testes A/B aplicados pela Netflix, ver https://bit.ly/3N62043. Acesso em: 9 mai. 2022.

31 Narrow AI. **DeepAI**. Disponível em: https://deepai.org/machine-learning-glossary-and-terms/narrow-ai.

32 IBM Cloud Education. O que é Inteligência Artificial? **IBM**, 3 de junho de 2020. Disponível em: https://ibm.co/3kXH3Oy.

Por último, a Superinteligência Artificial (SA) seria, em teoria, um estágio no qual a IA superaria a inteligência humana, sendo capaz de tomar decisões e armazenar dados com mais eficiência que os seres humanos. Além de executar tarefas e resolver problemas, uma SA poderia até mesmo demonstrar emoções e manter relacionamentos amorosos. Nick Bostrom define esse conceito como "qualquer intelecto que exceda a performance cognitiva de humanos virtualmente em todos os domínios de interesse".[33] Exemplos podem ser encontrados na ficção científica, como o computador HAL mencionado na introdução.

3. Princípios da IA

Além das concepções identificadas acima, o conceito de IA também é moldado e reconfigurado pela atividade de diferentes atores a partir de uma perspectiva multissetorial (seja governamental, empresarial, acadêmica ou com a participação de organizações da sociedade civil). Neste cenário, é possível identificar seis principais dimensões acerca dos princípios da IA que merecem destaque neste estudo, são elas: (i) Equidade; (ii) Confiabilidade e Segurança; (iii) Impacto Social; (iv) Responsabilidade; (v) Privacidade; e (vi) Transparência.[34]

A discussão a seguir será pautada, principalmente, pelas análises publicadas pela Comissão Europeia a respeito de padrões que garantem a confiabilidade da IA. Veja-se, entretanto, que a intenção não é oferecer um panorama completo do debate sobre os valores que devem permear o ecossistema da IA, mas sim ilustrar como esses princípios também constituem a própria ideia do que hoje entendemos por inteligência artificial. Afinal, os princípios representam a visão da nossa sociedade a respeito do futuro da IA e de seus impactos (positivos e negativos) para a humanidade, apontando caminhos para seu desenvolvimento saudável.

33 BOSTROM, Nick. **Superintelligence**: paths, dangers, strategies. Reino Unido: Oxford University Press, 2016.

34 BURLE, Caroline; CORTIZ, Diogo. Mapping Principles of Artificial Intelligence. **Ceweb.br**, 2019. Disponível em: https://bit.ly/3yrgMzT.

3.1 Equidade

O princípio da Equidade, talvez o mais importante entre os seis, é dividido em duas dimensões pela Comissão Europeia[35]: uma substantiva e outra procedimental. A primeira se reflete no compromisso de garantir a distribuição equitativa e justa de direitos e deveres, evitando preconceitos, discriminação e estigmatização, como também na promoção da igualdade de oportunidades no acesso a educação, bens, serviços e tecnologias.

Já a segunda dimensão, a procedimental, está associada à busca pela reparação efetiva frente às decisões tomadas por sistemas de IA e pelos humanos que as operam e desenvolvem, garantindo-se, assim, a identificação do responsável pela decisão e a explicabilidade dos processos de tomada de decisão às pessoas que interagem com aquela tecnologia, seja de forma direta ou indireta.

Em suma, como afirma a Organização para a Cooperação e Desenvolvimento Econômico (OCDE)[36], o princípio se reflete na necessidade de que os sistemas de IA sejam projetados respeitando o Estado de Direito, os direitos humanos, os valores democráticos e a diversidade, além do dever de se incluir salvaguardas necessárias para a garantia de uma sociedade equitativa e justa, promovendo a intervenção humana sempre que necessário para proteger esses valores.

3.2 Confiabilidade e Segurança

O princípio da Confiabilidade e Segurança, para a Comissão Europeia[37], estaria na condição de que sistemas de IA não devem cau-

35 Building Guidelines for Trustworthy AI European Commission. **European Commission**. Disponível em: https://bit.ly/3KT4dQA, pp. 12-13.

36 OECD. Principles on AI. **Organisation for Economic Co-operation and Development**. Disponível em: https://www.oecd.org/going-digital/ai/principles/.

37 Building Guidelines for Trustworthy AI European Commission. **European Commission**. Disponível em: https://bit.ly/3KT4dQA.

sar danos ou afetar adversamente os seres humanos, ao passo que os ambientes em que operam devem ser seguros, tecnicamente robustos e protegidos contra eventuais abusos e usos mal-intencionados. Além disso, deve ser despendida maior atenção às pessoas vulneráveis e grupos marginalizados, os quais precisam ser incluídos no desenvolvimento e implementação da IA com o intuito de mitigar impactos adversos em função de assimetrias de poder.

3.3 Impacto Social

O princípio do Impacto Social, para a Comissão Europeia[38], se traduz (i) na autodeterminação completa e eficaz dos seres humanos frente à IA, (ii) na proibição dos sistemas de IA subordinarem, coagirem, enganarem, manipularem, condicionarem ou agruparem pessoas de forma injustificada ou desproporcional, (iii) na alocação de funções entre humanos e sistemas de IA a partir do princípio de "design centrado no ser humano" (*human centered design*) e, por fim, (iv) na supervisão humana sobre os sistemas de IA.

3.4 Responsabilidade (*Accountability*)

Para a Comissão Europeia[39], o princípio da Responsabilidade engloba, dentre outros, mecanismos como auditorias independentes dos sistemas de IA, publicação de relatórios de impacto negativo pelos desenvolvedores dessas tecnologias e a criação de mecanismos para garantir a responsabilização e prestação de contas, seja antes, durante ou depois do desenvolvimento, implementação e uso de aplicações de IA no nosso dia a dia.

38 *Ibidem*, p. 19.

39 *Ibidem*, pp. 19-20.

3.5 Privacidade

A respeito do princípio da Privacidade, a Comissão Europeia[40] ressalta as seguintes dimensões: (i) respeito à privacidade e proteção de dados no contexto da IA, (ii) direito de acesso aos dados pelos seus titulares e (iii) prevenção de danos à privacidade, o que exige governança sobre a qualidade e integridade dos dados utilizados, sua relevância à luz do domínio em que o sistema de IA será implementado, seus protocolos de acesso e, por fim, seu processamento. Como se sabe, esses elementos constituem legislações paradigmáticas como o *General Data Protection Act* (GDPR) e a Lei Geral de Proteção de Dados (LGPD).

3.6 Transparência

Na visão da Comissão Europeia[41], o princípio da Transparência é crucial para manter a confiança dos usuários nos sistemas de IA. Para isso, (i) os processos e protocolos envolvidos nessas novas tecnologias devem ser transparentes; (ii) as capacidades e os objetivos dos sistemas de IA devem ser comunicados de forma clara e aberta; (iii) as decisões, na medida do possível, devem ser explicadas aos afetados direta e indiretamente; e (iv) em casos em que a explicabilidade não é possível – seja por motivos técnicos ou empresariais –, devem ser adotadas medidas alternativas como a rastreabilidade e auditabilidade das capacidades do sistema. Tudo isso, obviamente, ressalvados os segredos comerciais e industriais.

Conclusão: IA, um conceito em desenvolvimento

Em conclusão, a IA é um conceito dinâmico, aberto e, principalmente, em desenvolvimento. Ao nos questionarmos "o que é inteligên-

40 *Ibidem*, p. 17.

41 *Ibidem*, p. 18.

cia artificial?" é indispensável, em primeiro lugar, retomar a história do seu desenvolvimento desde que Alan Turing propôs um teste em 1950 para classificar máquinas como "pensantes" e John McCarthy organizou o primeiro evento acadêmico sobre o tema no verão de 1957 em Dartmouth. Não fossem algumas escolhas feitas pelos "pais fundadores" da IA, hoje poderíamos nos referir a termos como racionalidade computacional ou aprendizagem de máquina em seu lugar. Parte da resposta para o questionamento que dá título a este trabalho passa, portanto, por uma dependência de trajetória.

Em segundo lugar, também é importante organizar alguns subconceitos que estão englobados pelo universo da IA, o que nos permite avaliar exatamente o que essa tecnologia é e o que, ao menos em teoria, ela *pode ser*. Hoje a IA ainda está limitada ao desenvolvimento de tarefas específicas e à solução de problemas bem delimitados (IA Fraca). Aplicações que se assemelham à inteligência humana em uma perspectiva multidimensional (IA Forte) ou que até mesmo superam os seres humanos (Superinteligência Artificial) ainda só existem nos livros e filmes de ficção científica, o que não significa, entretanto, que não sejam possíveis em teoria e que não representam um potencial risco à humanidade.[42]

Por fim, em terceiro lugar, nossa visão compartilhada de futuro a respeito da IA também é parte constitutiva desse conceito essencialmente dinâmico. Os princípios que hoje são desenvolvidos e implementados por diferentes atores a partir de uma perspectiva multissetorial ajudam igualmente a responder à pergunta posta no início do capítulo. O que é inteligência artificial? É uma tecnologia equânime, confiável e segura, ciente do seu impacto social, responsável, protetora da privacidade e transparente. É impossível (e igualmente indesejável) separar o conceito de IA dos valores que elegemos para guiar o seu desenvolvimento e implementação.

42 BOSTROM, Nick. **Superintelligence**: paths, dangers, strategies. Reino Unido: Oxford University Press, 2016.

João Victor Archegas · Pesquisador Sênior do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio). Professor de Direito na FAE (Curitiba). Mestre em Direito (Master of Laws) pela Universidade Harvard. Gammon Fellow na Harvard Law School. Bacharel e mestrando em Direito pela Universidade Federal do Paraná (UFPR).

Gabriella Maia · Pesquisadora no V Grupo de Pesquisa do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio). Estagiária (Law Clerk) em Proteção de Dados no Tauil & Chequer Advogados associado a Mayer Brown. Membro da Comissão Nacional de Família e Tecnologia do Instituto Brasileiro de Direito de Família (IBDFAM). Graduanda em Direito pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).