Aoristo)))))
International Journal of Phenomenology, Hermeneutics and Metaphysics

Desmistificando a inteligência artificial: Uma breve introdução conceitual ao aprendizado de máquina

Demystifying artificial intelligence: A brief conceptual introduction to machine learning

Prof. Dr. Rafael Saraiva Campos
CEFET/RJ, Campus Petrópolis[1]

RESUMO

Este trabalho tem como objetivo apresentar os fundamentos da Inteligência Artificial (IA) para o público não-especializado. Dentro deste segmento, destacam-se os estudiosos da Filosofia, em particular da Filosofia da Mente, que têm em seu campo de estudo intersecções com a área da IA, quando debruçam-se sobre questões como inteligência e consciência, questões essas que são elementos centrais do trabalho de Hubert Dreyfus e John Searle. Conceitos fundamentais da IA - suas formas de atuação, aplicações e abordagens de implementação - são introduzidos de modo a desmistificar o tema. O trabalho traz também uma breve discussão sobre o problema mente-corpo e a consciência, com foco particular na abordagem de Searle, de modo a possibilitar a distinção entre os paradigmas da IA forte e fraca.

PALAVRAS-CHAVE

Inteligência; Aprendizado; Artificial; Consciência; Dreyfus; Searle.

ABSTRACT

This work presents the fundamentals of Artificial Intelligence (AI) to the non-specialized public. Within this group, those who study Philosophy stand out, particularly students of Philosophy of the Mind, whose studies intersect with AI related subjects, such as intelligence and conscience. Those are core themes within Hubert Dreyfus and John Searle works. Fundamental AI concepts - such as forms of

[1] Email: rafael.campos@cefet-rj.br. Orcid: https://orcid.org/0000-0001-9852-1362

acting, applications and implementation approaches - are introduced in order to demystify AI. This paper also brings a brief discussion of the body-mind problem and conscience, with particular focus on Searle's approach to the topic, in order to allow a proper distinction between strong and weak AI paradigms.

KEYWORDS
Intelligence; Learning; Artificial; Conscience; Dreyfus; Searle

INTRODUÇÃO

> Deve estar agora evidente a extrema dificuldade em definir-se o que seja o nível mental de funcionamento e que, seja a mente o que for, não é de modo algum óbvio que a mesma funcione como um computador digital. (DREYFUS, 1975, p. 157)

O objetivo primordial deste trabalho é desmistificar a inteligência artificial (IA) para o público não-especializado. Por público não-especializado referimo-nos àquele não treinado nos aspectos de desenvolvimento e implementação técnica de soluções nesta área, ou seja, o público em geral, especialistas e pesquisadores fora da Ciência da Computação e da Engenharia. Neste último grupo podem incluir-se pesquisadores acadêmicos de áreas como a Filosofia (em particular, a Filosofia da Mente) que queiram adquirir uma noção conceitual básica da IA e das técnicas de aprendizado de máquina.

O presente trabalho se justifica diante da popularização do tema dentre os mais diversos segmentos da sociedade. O interesse no assunto permeia grandes parcelas do público, diante do crescente ritmo de implementação de soluções de IA nas mais diversas áreas do conhecimento humano, desde o diagnóstico de doenças sem a necessidade de um médico até a navegação autônoma de veículos automotores. Contudo, este interesse reveste-se de uma preocupação subjacente, alimentada pela percepção dos possíveis desdobramentos deletérios da IA, dentro os quais destaca-se o desemprego. Além deste medo de substituição do homem no mercado de trabalho pela máquina inteligente[2], há um temor mais abrangente e profundo, de que o homem tornar-se-á obsoleto, e que a IA apresentar-se-á tal qual um novo estágio num processo evolutivo não mais puramente biológico, mas cibernético, obliterando a presença do homem em todas as esferas, levando-o à extinção (OPEN PHILANTROPY, 2015).

A falta de informação clara e despida de viés sensacionalista para o público não-especializado alimenta preocupações e medos, por vezes distópicos e desprovidos de

[2] A Seção 2 introduz uma definição de inteligência a ser considerada no contexto da IA.

qualquer fundamento técnico. Exatamente neste ponto crucial o presente trabalho se insere, apresentando de forma sintética os fundamentos técnicos da IA e do aprendizado de máquina, ao mesmo tempo em que busca auxílio de recursos da Filosofia para contextualizar conceitos como inteligência e consciência, conceitos estes imprescindíveis para a correta distinção entre os paradigmas da IA fraca e forte.

O restante do trabalho está dividido como se segue. A Seção 1 traz uma visão geral da IA, apresentando: grandes áreas e aplicações; as tecnologias subjacentes que possibilitam seu desenvolvimento e operação; sua característica multidisciplinar (onde modelos construídos por engenheiros apropriam-se de elementos fornecidos pela Matemática, Linguística, Neurociência e Biologia, dentre outras) e os agentes que atuam na construção da percepção pública sobre o tema; seus impactos socioeconômicos, com foco particular nas previsões de redução de oferta de vagas de trabalho em diferentes áreas. A Seção 2 traz um conceito de inteligência básico para o entendimento do que é a IA, além de descrever sucintamente os principais tipos de aprendizado de máquina. A Seção 3 apresenta as principais abordagens para a implementação de modelos de aprendizado de máquina. A Seção 4 faz a distinção entre a IA fraca e a IA forte. Para o discernimento entre esses dois paradigmas, é imprescindível abordar a questão da consciência e do problema mente-corpo, temas estes centrais para a Filosofia da Mente. Tal abordagem, é claro, será sucinta, dada a complexidade do tema.

1. VISÃO GERAL

1.1. GRANDES ÁREAS E APLICAÇÕES

> O campo da inteligência artificial tem muitas divisões e subdivisões, mas o trabalho mais importante pode ser classificado em quatro áreas: participação em jogos, tradução de idiomas, solução de problemas e reconhecimento de padrão. (DREYFUS, 1975, p. 38)

Dreyfus compartimentalizou a inteligência artificial em quatro grandes áreas: participação em jogos, tradução de idiomas, solução de problemas e reconhecimento de padrões. Com uma pequena modificação, podemos partir desta divisão para categorizar as aplicações de IA.

A tradução de idiomas já era alvo de aplicações de IA desde sua primeira fase, na época da IA simbólica no final dos anos 1950. Os êxitos no setor foram, e em certa medida ainda o são, limitados, sobretudo devido à dificuldade dos programas para lidar com a ambiguidade inerente à linguagem. Contudo, na categorização da IA, esta subdivisão deve ser expandida para englobar todas as aplicações envolvendo *processamento de linguagem natural*, i.e., conversão de fala para texto, interpretação de

comandos de voz, entre outros. Nesta categoria de aplicações de IA incluem-se os chamados *chatbots*, programas capazes de simular o discurso humano, interagindo com o usuário, na maioria das vezes por interface textual, mas também por meio de simuladores de voz (como nas centrais de atendimento automático).

Na *participação em jogos*, o primeiro exemplo foram as tentativas de criar um programa capaz de jogar xadrez. Na primeira fase da IA, no final dos anos 1950, foram desenvolvidos programas capazes de derrotar jogadores principiantes. O cenário modificou-se bastante nas décadas seguintes, e como é de conhecimento geral, um computador (o *Deep Blue* da IBM) foi capaz de vencer o campeão mundial de xadrez, já em 1997. Obviamente, atualmente, a atuação da IA não se restringe a jogos de tabuleiro, como o xadrez, dama ou o go, mas também se estende a jogos de videogame de tiro em primeira pessoa (FPS – *First Person Shooting*), como os famigerados *Doom* e *Duke Nukem*, onde o computador controla personagens do jogo (*bots*), os quais devem ser enfrentados pelos (ou atuar cooperativamente com os) jogadores humanos em ambientes de simulação tridimensional.

A área de *reconhecimento de padrões* envolve a identificação de padrões em textos, no discurso falado, e em imagens. Neste último caso, há o campo da *visão computacional*, onde o computador deve interpretar imagens, extraindo informações e significado das mesmas. A visão computacional é usada, por exemplo, na identificação automática de placas de automóveis a partir de imagens capturadas por câmeras de controle de tráfego. Outra aplicação da visão computacional é o reconhecimento facial. Ela também é essencial para o projeto de veículos autônomos. Outra aplicação que pode ser inclusa na categoria de reconhecimento de padrões seria a construção de conteúdo personalizado a partir da análise dos hábitos de consumo de um usuário de um serviço. Por exemplo, a montagem de listas personalizadas de músicas novas para o usuário, a partir da análise das músicas que o mesmo costuma escutar. Esta *personalização de conteúdo* é muito comum em aplicativos como o *Spotify*. Neste, quanto mais músicas o usuário escuta e avalia, tanto melhor o algoritmo consegue prever se o usuário vai gostar ou não de um lançamento musical.

Na categoria chamada por Dreyfus de *solução de problemas* podemos incluir aplicações de regressão não-linear e otimização, dentre outras. A regressão não-linear pode ser usada para a predição de séries temporais, por exemplo, na predição da demanda de um determinado medicamento, a partir do consumo deste medicamento nos meses anteriores. Já nos problemas de otimização, busca-se uma solução que maximize a função objetivo descrevendo a situação que se deseja otimizar.

1.2. TECNOLOGIAS SUBJECENTES

Na quase totalidade das situações, as tecnologias subjacentes às soluções de IA, que viabilizam seu desenvolvimento e operação, passam completamente

despercebidas do público-alvo. O conhecimento de tais tecnologias, embora não seja pré-requisito para uso das mesmas, é desejável àqueles que buscam uma compreensão mais profunda sobre o tema. De modo simplificado, tais tecnologias de suporte podem ser agrupadas em três classes: bancos de dados, tecnologias de acesso a dados, velocidade de processamento. A seguir discorremos de forma breve sobre cada uma delas.

Bancos de dados: as soluções de IA, implementadas por meio de algoritmos de aprendizado de máquina, requerem grandes massas de dados para aprender[3]. Via de regra, quanto mais complexo o modelo - sendo uma possível métrica desta complexidade a quantidade de parâmetros ajustáveis (livres) no modelo, como, por exemplo, os pesos sinápticos em uma rede neural artificial[4] - maior a quantidade de dados necessários para treiná-lo. Capacidade de armazenamento na forma digital certamente não é um fator limitante atualmente, com a disponibilidade de discos rígidos com dezenas de *terabytes*[5]. Além disso, por meio da difusão do acesso à Internet através de dispositivos móveis (principalmente *smartphones*), e através da popularização (que hoje assume uma forma praticamente onipresente) de plataformas de mídias sociais e de mecanismos de buscas *on-line*, bases de dados massivas são construídas de forma colaborativa e distribuída (através dos dados fornecidos pelos usuários destes dispositivos e plataformas). Portanto, há grande disponibilidade de dados na forma digital para o treinamento de modelos de aprendizado de máquina de elevada complexidade.

Tecnologias de acesso a dados: a enorme capacidade de armazenamento disponível atualmente não seria de grande serventia, a menos que fosse possível coletar tais dados de forma rápida e distribuída. Conforme já mencionado, grande parte desses bancos de dados recebem informações fornecidas por meio de dispositivos móveis. Desta forma, tecnologias de acesso sem fio via radiofrequência são cruciais para a construção e uso destas bases de dados. Dentre essas tecnologias de acesso, destacam-se os padrões de telefonia móvel celular de Quarta Geração (4G), que viabilizam taxas de transmissão de pico de até 1 Gbps (1 bilhão de bits por segundo)[6], e as redes locais sem fio com o padrão WiFi, hoje onipresentes em áreas urbanas. Além destas, há o acesso via Bluetooth[7]. Importante destacar que os *smartphones* atuais congregam todas estas

[3] A Seção 2 define o que é aprendizado, além de discorrer sobre as principais formas de aprendizado de máquina.

[4] Pesos sinápticos são parâmetros livres em uma rede neural artificial (RNA), que são modificados durante o aprendizado da rede. A RNA é um modelo matemático inspirado nas conexões sinápticas do cérebro. Os pesos sinápticos emulam matematicamente os potenciais de ativação destas sinapses neuronais. Para maiores detalhes, consultar a Seção 3.1.

[5] Um *terabyte* é igual a 10^{12} *bytes*, ou seja, 1 trilhão de *bytes*. Um *byte* corresponde a 8 bits, onde um bit é a unidade elementar de informação binária em computadores digitais, podendo assumir os valores um ou zero.

[6] Um bit é a unidade elementar de informação binária em computadores digitais, podendo assumir os valores um ou zero.

[7] Bluetooth é o protocolo padrão de comunicação para as chamadas redes de área pessoal (PAN – *personal area network*), com baixo consumo de energia e curto alcance (na maioria dos casos, apenas alguns metros) baseado em microchips transmissores de baixo custo em cada dispositivo.

tecnologias de acesso, usando-as de forma intercambiável. Por fim, deve-se ressaltar também os sistemas de satélite de navegação global, como o famigerado GPS (*Global Positioning System* - Sistema de Posicionamento Global), além dos menos conhecidos GLONASS da Rússia, o GALILEO da União Europeia e o COMPASS da China (CAMPOS; LOVISOLO, 2015). Tais redes de satélites são muito importantes para a construção de bases de dados com informações georreferenciadas, i.e., associadas a localizações geográficas[8].

Velocidade de processamento: a existência de grandes bases de dados, assim como a possibilidade de acesso rápido e ubíquo à informação armazenada não seriam de grande utilidade, se não fosse possível manipulá-las de forma rápida no processo de tomada de decisão autônoma pela IA. Essa necessidade de rapidez é facilmente ilustrada no caso de navegação autônoma de veículos automotores, ou seja, carros sem motoristas. Grande quantidade de dados é coletada pelos diversos sensores de tais veículos - imagens por meio de câmeras, estimativas de distâncias a obstáculos por meio de sensores acústicos ou radares, dados de tráfego coletados da Internet via redes de telefonia móvel celular para o traçado otimizado de rotas, dentre outros. Tais dados devem ser processados (interpretados) de forma rápida. Em certos casos o tempo para a tomada de decisão é de milissegundos (como, por exemplo, quando outro veículo subitamente avança o sinal vermelho em um cruzamento logo à frente do carro autônomo pilotado pela IA). Em situações como esta, uma elevada capacidade de processamento (sinônimo de velocidade) é essencial. Uma métrica tipicamente empregada para quantificar a velocidade de processamento é o *flops* (*floating point operations per secon*d - operações de ponto flutuante[9] por segundo). Por exemplo, o processador Intel i486DX4 100 MHz lançado no início dos anos 1990 tinha uma capacidade de 7 mega FLOPS, ou seja, 7 milhões de operações de ponto flutuante por segundo. Processadores da classe INTEL Xeon, como o E5 2699 v4, lançado em 2016, atingem taxas de 2 tera FLOPS, ou seja, 2 trilhões de operações de ponto flutuante por segundo. Isto corresponde a um aumento de aproximadamente 28 mil vezes na velocidade de processamento em menos de 30 anos. Placas de processamento gráfico (GPUs - *Graphical Processing Units*) como a NVDIA GeForce Titan X já atingiam taxas de 10 tera FLOPS em 2016, o que corresponde a um aumento de cerca de 140 mil vezes em relação aos processadores dos anos 1990 (RUPP, 2016). Dentre os computadores de grande porte (os chamados supercomputadores), que utilizam múltiplos processadores em paralelo, o mais poderoso em operação atualmente é o IBM *Summit*, com uma velocidade de processamento de pico de 200 peta FLOPS, ou seja, 200000

[8] Um exemplo de *software* que faz uso de informações georreferenciadas é o aplicativo de transporte *Uber*, onde a localização do usuário e do motorista são continuamente registradas, desde a solicitação da corrida até seu encerramento.

[9] A representação de números na forma binária em ponto flutuante segue o seguinte padrão: o sinal (positivo ou negativo) precedendo a mantissa (parte fracionária), que por sua vez multiplica a base 2 elevada a uma potência E. A representação em ponto flutuante é normalizada pelo padrão IEEE 754 (IEEE, 2008).

trilhões de operações de ponto flutuante por segundo (IBM, 2018). Esta velocidade de processamento, ainda que impressionante, será eclipsada quando do advento dos computadores quânticos, ainda em desenvolvimento (YING, 2010). Taxas de processamento de informação tão elevadas significam que modelos de aprendizado de máquina de alta complexidade poderão manipular quantidades gigantescas de informação, identificando padrões e tendências que servirão de suporte a processos decisórios autônomos baseados em IA.

1.3. CONSTRUÇÃO DE PERCEPÇÃO PÚBLICA

As estratégias de comunicação embasadas em um modelo linear de transferência de conhecimento do tipo especialista-leigo reconhecidamente fracassaram na tentativa de aproximar a sociedade em geral da ciência (SABBATINI, 2004). Assim, com relação à ciência, e particularmente no que concerne à IA, a construção da percepção pública alicerça-se essencialmente sobre dois pilares não-especializados: a imprensa e o cinema.

A imprensa, muitas vezes pela pressão por audiência, usa de manchetes bombásticas para atrair a atenção do público. Por exemplo, um recente artigo da BBC (British Broadcasting Corporation) foi intitulado "Como a inteligência artificial poderia acabar com a Humanidade - por acidente" (BBC, 2019). Outro exemplo é um recente artigo da Revista Época Negócios, intitulado "Como a inteligência artificial vai sair do controle e dominar o mundo" (GIL, 2019). Muitas vezes, tais artigos com títulos alarmistas e sensacionalistas contribuem para a desinformação e confusão, ao invés de esclarecer efetivamente as dúvidas sobre o tema da IA e seus possíveis impactos sobre a sociedade.

Outro agente crucial na formação da opinião pública sobre a IA é o cinema. Este, ao contrário da imprensa, não tem a obrigação precípua de informar, mas apenas de entreter. O cinema é responsável pela criação do imaginário popular em torno da IA através de filmes revolucionários para suas épocas, como *2001*, de Stanley Kubrik, lançado em 1968, e *Terminator* 1 e 2, lançados em 1984 e 1991, respectivamente. Outros títulos dignos de nota são *Matrix*, de 1999, e *Blade Runner*, de 1982, além do clássico *Metrópolis*, de 1927, provavelmente o mais antigo filme abordando o tema das "máquinas inteligentes". Todos os filmes aqui citados têm uma abordagem pessimista, onde a IA age de forma autônoma para extirpar a humanidade, na qual vê uma séria ameaça, ou onde a IA é uma ferramenta de manipulação do homem pelo homem.

1.4. IMPACTOS SOCIOECONÔMICOS

Por trás dos temores a respeito de um futuro distópico dominado por máquinas que percebem no homem uma forma inferior de existência destinada a ser extirpada - temores estes alimentados pelo cinema (que cremos ser inimputável, pois goza da

licença poética) - há uma ameaça muito mais palpável e iminente representada pela IA: o desemprego.

Embora aqueles que exercem profissões técnicas que exigem alto grau de qualificação, como médicos, engenheiros e advogados, possam crer-se imunes a esta ameaça, muitas aplicações já existentes da IA, bem como diversas projeções para o futuro próximo, indicam que eles estão enganados. Aplicações de visão computacional para o diagnóstico de doenças por imagens já atingem taxas de acerto superiores a médicos (WALSH, 2020). Programas de processamento de linguagem natural são capazes de elaborar petições legais, e até mesmo prever o resultado de julgamentos, bem como as sentenças impostas, a partir da análise de documentos (MANGAN, 2018). Aplicações de IA para a otimização de redes de acesso rádio (como as redes de telefonia móvel celular) são capazes de substituir a análise de engenheiros de telecomunicações, propondo ações para a melhoria da qualidade da rede (CAMPOS; LOVISOLO, 2019).

De fato, uma pesquisa recentemente publicada pela FGV, em conjunto com a Microsoft, analisou o impacto da IA em seis setores da economia brasileira: agricultura, pecuária, óleo e gás, mineração e extração, transporte e comércio e setor público (educação, saúde, defesa e administração pública). A pesquisa prevê que os trabalhadores mais gravemente afetados pela IA serão os mais qualificados nos setores de óleo e gás e de agricultura, com redução de 23.57% e 21.55% na oferta de vagas, respectivamente, nos próximos 15 anos (FGV, 2019).

Em um outro estudo, pesquisadores da Universidade de Oxford desenvolveram uma metodologia para prever a redução das vagas causadas pela IA em diferentes segmentos do mercado de trabalho (FREY; OSBORNE, 2013). Por exemplo, este estudo estima que a redução das vagas nos próximos 20 anos nas profissões digitador e motorista será de 91% e 88%, respectivamente. Ou seja, haverá uma automatização quase completa de tais atividades.

Contudo, embora haja muitas previsões desencorajadoras, elas não são universalmente aceitas como o único desfecho. Há estudos que consideram que novas profissões serão geradas, em substituição àquelas extintas pela automação, e que os salários e a qualidade de vida em geral melhorarão, dependendo da velocidade e magnitude do desenvolvimento e difusão das tecnologias de IA, um ponto sobre o qual os especialistas divergem amplamente (MIAILHE; HODES, 2017).

2. INTELIGÊNCIA E APRENDIZADO

2.1. INTELIGÊNCIA

O uso do adjetivo "inteligente" para qualificar a atuação de um programa de computador que joga xadrez, por exemplo, causa estranheza e repúdio em muitas

pessoas. De modo geral, assume-se que a Inteligência é um atributo exclusivamente humano, e que envolve habilidades multifacetadas, tanto do ponto de vista racional-lógico quanto emocional.

De fato, há uma enormidade de definições para o que seria a Inteligência. Todavia, aqui, apenas uma definição – bastante prosaica e direta – nos interessa. E esta definição é à qual deveremos recorrer todas as vezes que nos depararmos com o termo "Inteligência Artificial". A definição em questão é: *Inteligência é a capacidade de aprender.* Qualquer programa de computador que seja capaz de aprender – desde a jogar xadrez até reconhecer rostos ou pilotar um avião – é dito inteligente, sob esta ótica. Mas, uma pergunta então se apresenta de imediato: *o que é aprender*?

2.2. APRENDIZADO

O que significa aprender algo? O que é aprendizado? Este é um conceito tão pervasivo na Natureza que raramente debruçamo-nos sobre ele. Aprendizado pode ser definido como o processo de familiarização com uma determinada situação, de modo a tornar-se apto a reagir adequadamente para atingir um objetivo específico. Na Natureza, "tornar-se apto a reagir adequadamente" a uma certa situação, muitas vezes, significa a diferença entre a sobrevivência e a morte. Por "situação" referimo-nos a eventos, condições ou informação (dados, ou, na terminologia do aprendizado de máquina, padrões de entrada). Há três pilares sobre os quais sustenta-se o processo de aprendizado: experiência, memória e generalização.

Indivíduos de uma espécie adquirem novas habilidades (i.e., eles aprendem) através da experiência: eles são repetidamente expostos a uma determinada situação até que sejam aptos a reagir adequadamente quando confrontados com a mesma situação no futuro. Obviamente, a experiência seria inútil se não pudesse ser preservada na memória. A experiência deve ser acumulada, e sem a memória ela seria completamente perdida. Todavia, há um problema em recorrer unicamente à experiência, já que raramente uma situação ocorre exatamente da mesma forma mais de uma vez. Assim, experiência, sem a capacidade de generalizar, também não seria de grande valia. Por generalização referimo-nos à habilidade de extrapolar a experiência acumulada, aplicando-a a uma situação nova. Ou seja, é a capacidade de identificar similaridades entre novas circunstâncias e aquelas previamente encontradas[10].

[10] Para maior esclarecimento, considere o seguinte exemplo: um estudante de Matemática resolve centenas de exercícios em sala de aula, acumulando experiência em sua memória para preparar-se para o exame. Durante a prova, o estudante terá que elucidar problemas de Matemática que provavelmente nunca viu (usualmente professores não repetem na prova os exercícios resolvidos em sala). O estudante deverá então ser capaz de generalizar a experiência adquirida para conseguir resolver as questões da prova.

2.3. APRENDIZADO DE MÁQUINA

Neste contexto, o aprendizado de máquina é o processo de aprendizado aplicado a computadores, mais especificamente, a programas de computador (*software*). Todos os conceitos de aprendizado a partir da experiência, memória e generalização são tomados da Natureza e aplicados às máquinas. Todavia, no aprendizado de máquina, tudo reduz-se a números.

Há essencialmente quatro tipos de aprendizado de máquina: supervisionado, não-supervisionado, evolutivo e por reforço. Discorreremos sobre os dois primeiros a seguir. O terceiro tipo será tratado na Seção 3.3. O quarto tipo foge do escopo deste texto.

Aprendizado Supervisionado: neste tipo de aprendizado, a máquina recebe padrões de entrada para os quais as saídas esperadas (também denominadas alvos) são conhecidas[11]. A máquina então ajusta seus parâmetros internos de modo a atingir o resultado esperado. Uma função de custo (que deve ser minimizada) é usada para guiar estes ajustes. Ou seja, o aprendizado da máquina ocorre por meio da modificação de seus parâmetros internos[12]. Ao final da fase de aprendizado ou treinamento, a máquina deverá ser capaz de generalizar, produzindo as saídas esperadas para padrões de entrada semelhantes, porém não idênticos àqueles fornecidos durante a fase de aprendizado (treinamento). Note que os três pilares do aprendizado estão aqui inseridos: experiência, por meio dos padrões de entrada fornecidos durante o treinamento; memória, através da preservação dos parâmetros internos da máquina em seu estado final após o treinamento; e a generalização, obtida de diferentes formas, dependendo do tipo de máquina (ver Seção 3).

Aprendizado Não-Supervisionado: em alguns casos, não se conhecem os alvos (saídas esperadas) para os padrões de entrada fornecidos durante o treinamento. Nesses casos, o treinamento supervisionado obviamente não é viável. Em tais circunstâncias, a máquina, ao invés de produzir uma saída esperada (o que é impossível, diante da inexistência de alvos), deve ser capaz de identificar similaridades entre os padrões de treinamento, agrupando-os em conjuntos (classes). Esta similaridade é definida matematicamente, por meio de uma métrica adequada ao problema.

[11] Por exemplo, os padrões de entrada podem ser as características físicas de uma imagem tomográfica de um pulmão. Os alvos podem ser as classificações das imagens como pulmões saudáveis ou não.

[12] O que se entende por "parâmetros internos" depende do tipo de "máquina". Por exemplo, no caso de uma rede neural artificial (ver Seção 3.2), os parâmetros internos modificados durante a aprendizagem supervisionada são os pesos sinápticos da rede.

3. ABORDAGENS

Há diferentes formas de implementar modelos de aprendizado de máquina. Essas abordagens, contudo, podem ser agrupadas em três classes: a abordagem simbólica, a abordagem conexionista e a abordagem evolutiva. A primeira baseia-se num pressuposto epistemológico a respeito da Lógica Clássica, enquanto as duas outras alicerçam-se sobre pressupostos inspirados na Biologia.

3.1. SIMBÓLICA

A abordagem simbólica foi a base da primeira geração da IA (1957-1962), uma época onde os estudos na área focavam na simulação cognitiva. Neste período, havia um grande otimismo (que se revelou precipitado) em relação ao potencial da IA de replicar a cognição humana. Boa parte deste otimismo devia-se ao chamado pressuposto epistemológico. Segundo Dreyfus,

> [...] o pressuposto de que o homem funciona tal qual um mecanismo genérico de manipulação de símbolos implica [...] um pressuposto epistemológico de que todo conhecimento pode ser formalizado, isto é, que tudo que possa ser compreendido pode ser expresso em termos de relações lógicas. (DREYFUS, 1975, p.120)

Contudo, este pressuposto não é axiomático. Nas palavras do próprio Dreyfus:

> A afirmação [...] de que é possível formalizar todo o comportamento não arbitrário, não constitui um axioma, sendo antes a expressão de uma certa concepção de entendimento profundamente arraigada em nossa cultura, mas que ainda assim pode revelar-se errônea. (DREYFUS, 1975, p.159)

A abordagem simbólica faz uso de uma representação declarativa do conhecimento, i.e., usa a lógica proposicional para construir conjuntos de regras axiomáticas. O exemplo mais comum de implementação de IA segundo a abordagem simbólica são os chamados sistemas especialistas[13]. Nestes, uma base de conhecimento é construída por meio da declaração de fatos. Há também um motor de inferência, que acessa a base de conhecimento e faz uso de um conjunto de regras lógicas para decidir a decisão a tomar.

[13] Sistemas especialistas são muito usados para diagnósticos de enfermidades. Um exemplo é o sistema especialista usado para o diagnóstico de policitemia vera, uma doença hematológica (ANES; FORTES, 2019).

3.2. CONEXIONISTA

A abordagem conexionista, ou conexionismo, baseia-se na aprendizagem (supervisionada ou não), dispensando a necessidade de representação declarativa do conhecimento por meio da lógica ou da linguagem natural (BITTENCOURT, 2006).

O modelo matemático usado na abordagem conexionista é a chamada rede neural artificial (RNA). RNAs são sistemas distribuídos paralelos compostos por unidades de processamento interconectadas chamadas neurônios, que realizam operações matemáticas (denominadas funções de transferência ou de ativação). As conexões entre os neurônios são denominadas sinapses, e a cada uma é atribuído um peso numérico. Cada neurônio tem *n* entradas. Analogamente aos dendritos em um neurônio no cérebro humano, estas entradas recebem estímulos (valores numéricos) das sinapses com neurônios vizinhos. Estas entradas são agregadas (combinadas linearmente). A este valor agregado é adicionado um potencial de ativação, e por fim o valor resultante é fornecido à função de ativação do neurônio. O valor de saída é passado adiante, para os neurônios seguintes. A saída do neurônio na RNA emula o axônio de um neurônio humano.

O treinamento (aprendizado) de uma RNA é inspirado pelo processo de aprendizado e armazenamento de memórias no cérebro humano: as conexões (sinapses) que são mais frequentemente ativadas tem seus pesos aumentados (conexões sinápticas mais fortes, i.e., disparadas mesmo com estímulos mais fracos), enquanto conexões pouco usadas tem seus pesos reduzidos (eventualmente caindo no "esquecimento", já que a informação que representam não é usada). Este processo baseia-se na regra verificada por Hebb em neurônios humanos, que indica que as mudanças na força das conexões sinápticas são proporcionais à correlação da ativação de dois neurônios conectados (MARSLAND, 2009).

A abordagem conexionista encontra sua expressão mais poderosa nas chamadas redes neurais de aprendizado profundo (DNN – *Deep Learning Neural Networks*), que têm obtido resultados excelentes em áreas como visão computacional, reconhecimento de fala e processamento de linguagem natural. As DNNs são modelos conexionistas com número muito elevado de neurônios (às vezes, dezenas de milhões), o que as permite aprender a reconhecer padrões bastante complexos, como, por exemplo, a identificação automática de armas a partir de imagens de vídeo em tempo real, possibilitando a ativação de alarmes sem necessidade de intervenção humana, ou seja, dispensando a presença de funcionários monitorando continuamente as imagens das câmeras de segurança (OLMOS; TABIK; HERRERA, 2018).

3.3. EVOLUTIVA

O paradigma da abordagem evolutiva baseia-se na teoria da seleção natural e evolução das espécies de Charles Darwin. Indivíduos de uma espécie são definidos

por seu código DNA (sequência de genes nos cromossomos). Os genes são transmitidos de geração em geração. Indivíduos mais aptos (dentro do contexto ambiental em que a espécie existe) tem maior chance de sobreviver e reproduzir. Reprodução sexuada (com troca de material genético entre dois indivíduos) e a mutação (alteração aleatória dos genes que comumente resultam em indivíduos não-viáveis, porém que ocasionalmente propiciam uma vantagem adaptativa) promovem variabilidade genética, permitindo o surgimento de novos indivíduos mais aptos (evolução da espécie). Este processo de seleção/evolução pode ser encarado como uma forma de aprendizado coletivo, em que o DNA representa a memória do aprendizado acumulado de uma espécie, e é a essência da abordagem evolutiva, largamente aplicada a problemas de otimização em Engenharia (CAMPOS; LOVISOLO, 2019).

Em problemas de otimização, busca-se a solução que maximiza a função objetivo que descreve o problema (por exemplo, qual a sequência de passos na montagem de um automóvel em uma linha de produção maximiza o número de carros montados por hora?). Contudo, particularmente em problemas de otimização multivariados, o espaço de soluções possíveis é tão grande que resolver o problema por tentativa e erro (testar todas as soluções possíveis, selecionando a melhor) é computacionalmente inviável. Nestes casos, os chamados *algoritmos genéticos* fornecem um mecanismo para reduzir o tempo de busca dentro do espaço de soluções. A busca é guiada por uma modelagem matemática do processo de seleção natural e evolução genética, descrito no parágrafo anterior.

Os algoritmos genéticos executam um ciclo evolutivo (cada etapa do ciclo é chamada de *geração*), cujos passos podem ser resumidos como se segue:

i. Inicialmente, possíveis soluções para o problema são aleatoriamente selecionadas. Na nomenclatura dos algoritmos genéticos, cada solução candidata é um *indivíduo*. O conjunto de indivíduos inicialmente selecionado é a *população inicial*. Cada indivíduo é representado através de uma sequência numérica chamada *cromossomo*. Os diferentes atributos de uma solução candidata são representados numericamente em seções (*alelos*) do cromossomo.
ii. A aptidão de cada indivíduo é calculada utilizando uma *função de avaliação*, que é a *função objetivo* do problema de otimização. O objetivo em um problema de otimização é encontrar o indivíduo mais apto, i.e., aquele que maximiza a função objetivo.
iii. De posse das aptidões de cada indivíduo, um mecanismo de *seleção* para reprodução é aplicado. Assim como no ambiente natural, indivíduos mais aptos tem maiores chances de sobreviver e reproduzir. O conjunto de indivíduos da população atual selecionado para reprodução constitui o *mating pool*.
iv. Os indivíduos do *mating pool* são pareados e seus cromossomos seccionados em um ou mais pontos. Estas seções de cromossomos dos dois indivíduos são misturadas, simulando a recombinação genética, ou seja, a troca de material genético entre os cromossomos de dois indivíduos na reprodução sexuada, processo este denominado *cross-over*. Assim, cada par de indivíduos da geração

atual produzirá dois novos indivíduos para a próxima geração. Ao longo das gerações, a aptidão média da população aumenta, ou seja, melhores soluções para o problema de otimização vão sendo encontradas.

v. *Mutação* é aplicada sobre os cromossomos dos novos indivíduos. A mutação é uma alteração aleatória dos genes, que comumente resulta em indivíduos inviáveis. Todavia, em algumas situações, a mutação pode produzir indivíduos mais aptos. No contexto de algoritmos genéticos, o *cross-over* e a *mutação* permitem explorar mais soluções dentro do espaço de busca (conjunto de soluções possíveis).

O ciclo evolutivo descrito acima continua até que um critério de parada seja atingido (por exemplo, número máximo de gerações). A solução do problema é dada pelo indivíduo (i.e., pela solução candidata) mais apto na última geração.

Algoritmos genéticos são uma meta-heurística, i.e., um procedimento que seja capaz de encontrar uma solução suficientemente boa para um problema. Tal solução é dita sub-ótima. De fato, como uma meta-heurística toma amostras (e as avalia) apenas de um subconjunto das soluções possíveis (em situações em que o conjunto de todas as soluções é demasiadamente grande para ser avaliado por completo), não se pode assegurar que a solução encontrada seja a melhor de todas as soluções. Essa é uma das principais críticas contra o uso da abordagem evolutiva. Todavia, é inegável que esta abordagem – particularmente através de sua principal expressão, os algoritmos genéticos – tem sido usada como sucesso em ampla variedade de problemas.

4. IA FRACA VERSUS IA FORTE

> [...] se fôssemos produzir consciência artificialmente, a maneira natural de agir seria tentar reproduzir o fundamento neurobiológico efetivo que tem a consciência em organismos como nos próprios. Porque atualmente não sabemos exatamente qual é esse fundamento neurobiológico, as perspectivas para tal inteligência artificial são muito remotas. (SEARLE, 2006, p. 136)

A IA Forte é a hipótese de que as máquinas podem adquirir consciência, tal qual a humana. A IA Fraca é a hipótese de que as máquinas podem exibir comportamento inteligente indistinguível do humano, mas que não são verdadeiramente conscientes. Ou seja, na hipótese da IA Fraca, admite-se apenas que uma simulação de consciência é viável, mas que a consciência de fato não é alcançável para uma máquina. Aqui, defrontamo-nos com questões essenciais, e ainda sem respostas definitivas. O que é consciência? A consciência é verificável externamente? Ou seja, ela é uma experiência comunicável? Ou é inescrutável, dado que não é possível a transposição entre sinais

cerebrais (coletados por um tomógrafo ou um eletroencefalograma, ou por qualquer outro dispositivo capaz de monitorar a atividade cerebral) e conteúdos mentais (AIUB, 2009)?

Pode-se definir consciência, de modo bastante simples, como uma experimentação subjetiva da realidade. Com base nesta definição, surge a questão: uma entidade que exibe comportamento inteligente, indistinguível daquele apresentado por uma entidade assumidamente consciente (nós, humanos) é consciente? Ou seja, uma máquina hipotética que fosse capaz de emular à perfeição o comportamento humano nas mais diversas e complexas situações, seria de fato consciente? Não há uma resposta única para esta pergunta. Uma resposta possível seria que, do mesmo modo que assumimos que as pessoas são conscientes mesmo não tendo acesso direto aos estados mentais internos de outros seres humanos, devemos assumir a convenção de que máquinas que exibem comportamento inteligente de fato pensam (RUSSEL,2013). Outra resposta é que dois sistemas podem ter comportamentos indistintos, sendo um consciente e outro completamente inconsciente (desprovido de um sujeito que experimenta a realidade), logo a simulação de consciência não equivale à consciência (BLACKMORE, 2012).

Em relação à consciência, as perguntas são abundantes e as respostas, elusivas. Porém, no que concerne à IA, todas as implementações existentes e em desenvolvimento são exemplos de IA Fraca. O que há hoje é uma miríade de implementações de IA que atuam como especialistas em áreas muito específicas – por exemplo, um programa para emitir diagnósticos de tumores de pulmão a partir de imagens tomográficas, um programa para prever a cotação de um determinado ativo na bolsa de valores, um programa para prever a demanda por determinado medicamento no mercado internacional, um programa para jogar xadrez, e assim por diante. Ou seja, as implementações são muito eficientes – superando, em muitos casos, os especialistas humanos, como no caso do jogo de xadrez – em campos muito restritos de aplicação. Elas são incapazes de atuar de forma genérica, como um humano adulto, diante de problemas quaisquer. Ou seja, falta à IA na forma atual uma capacidade ampla de generalização (um dos pilares da definição de inteligência). Há esforços em andamento para o desenvolvimento de uma IA de propósito geral[14], i.e., uma IA capaz de aprender uma ampla gama de tarefas em áreas distintas (LONG, 2019). Contudo,

[14] Para melhor ilustração, considere o seguinte exemplo. Assuma que temos um programa de IA restrita, capaz de aprender a jogar xadrez, superando até mesmo o melhor jogador (humano) do mundo. Este programa presta-se tão somente a jogar xadrez, e nada mais. Considere agora que uma nova versão deste programa é capaz de aprender a jogar outros jogos de tabuleiro, além de xadrez, e também a utilizar este conhecimento como suporte auxiliar ao processamento de linguagem natural, de textos sobre xadrez ou outros jogos de tabuleiro que tenha aprendido. Neste caso, o programa é capaz de realizar múltiplas atividades distintas – jogar diversos jogos de tabuleiro e analisar textos – e, além disso, é capaz de transferir conhecimento de uma atividade para a outra. Este seria um exemplo rudimentar de IA de propósito geral (GPAI – *General Purpose Artificial Intelligence*). Na sua versão plena, um programa com suporte a GPAI seria capaz de executar grande variedade de tarefas, transferir conhecimentos entre as bases de conhecimentos usadas para a realização de cada tarefa, além de poder aprender tarefas novas, não inicialmente previstas quando o programa foi escrito.

não há um projeto de IA Forte[15], ou seja, não há por parte dos desenvolvedores uma aspiração por criar consciência verdadeira nas máquinas. Isto se dá por duas razões essenciais: a primeira é que é inconsistente almejar replicar algo para o qual não se tem sequer uma definição fechada (a consciência); a segunda é que a consciência não é pré-requisito para que a IA possa operar no mundo (como atestam os incontáveis exemplos de aplicações da IA Fraca existentes).

Particularmente, partilhamos da opinião de Searle, que pode ser sintetizada em sua frase "cérebros causam consciência" (SEARLE, 1992, p. 137). Portanto, enquanto não forem compreendidos de forma clara os processos bioquímicos que geram a consciência – ou seja, enquanto não for entendida como se dá a emergência psicofísica, i.e., o surgimento da mente a partir de processos físicos no cérebro – será impossível replicar tal processo em máquinas. Sem este entendimento, o projeto da IA Forte é inexequível.

CONCLUSÃO

Este artigo apresentou uma visão geral da IA apontando quais seus pilares tecnológicos fundamentais, os agentes envolvidos na construção da percepção pública sobre o tema e os impactos esperados na sociedade, particularmente na oferta de empregos. Em seguida introduziu conceitos de inteligência e aprendizado no contexto da IA discorrendo sobre os tipos principais de aprendizado de máquina. Por fim, abordou a questão da consciência, a fim de possibilitar a distinção entre os paradigmas da IA fraca e forte.

O artigo tinha como objetivo primordial desmistificar o tema IA. Neste contexto, dois itens devem ser ressaltados na conclusão, no que concerne ao estágio atual da IA, e também aos desenvolvimentos que podemos discernir no futuro próximo. Primeiro, hoje dispomos de uma miríade de implementações de IA que são altamente especializadas para a solução de problemas muito específicos, inexistindo uma implementação de aplicação geral, sendo este o objetivo da área de pesquisa denominada Aprendizado de Máquina Automático (*AML – Automated Machine Learning*) que ainda se encontra em estágios iniciais. Segundo, a indagação sobre se as máquinas podem adquirir consciência mostra-se precipitada, já que, como já defendia Searle, não é possível reproduzir consciência enquanto os processos neuro-físicos que promovem a emergência psico-física no cérebro humano não forem completamente compreendidos.

[15] Esta afirmação baseia-se, dentre outras coisas, na análise dos anais de 1994 a 2018 do renomado Congresso Mundial de Inteligência Computacional (WCCI – *World Congress on Computational Intelligence*), organizado pelo IEEE (Instituto de Engenheiros Elétricos e Eletrônicos). No período de 27 anos considerado, não há qualquer trabalho referente à IA Forte.

REFERÊNCIAS

AIUB, M. *Filosofia da Mente e Psicoterapias.* Rio de Janeiro: Wak, 2009.
ANES, L. F e FORTES, R. S*. Sistema Especialista para o auxílio no Diagnóstico da Policitemia Vera*. Barbacena, MG (Trabalho de Conclusão de Curso). Universidade Presidente Antônio Carlos – UNIPAC, 2019.
BBC. Como a inteligência artificial poderia acabar com a Humanidade - por acidente, Outubro 2019 <bbc.com/portuguese/geral-50228913> Acessado em 1 de Fevereiro de 2020.
BITTENCOURT, G. *Inteligência Artificial*: Ferramentas e Teorias. 3ed. Florianópolis: Editora da UFSC, 2006.
BLACKMORE, S. *Consciousness*: An Introduction. 2ed. New York: Oxford University Press, 2012.
CAMPOS, R. S.; LOVISOLO, L. *RF Positioning: Fundamentals, Applications and Tools*. Boston: Artech House, 2015.
____. Genetic algorithm-based cellular network optimisation considering positioning applications. In: *IET Communications*. Vol 13(7), 2019, p. 879-891.
DREYFUS, H. L. *O que os Computadores não podem Fazer* - Uma Crítica da Razão Artificial. Rio de Janeiro: A Casa do Livro Eldorado, 1975.
FREY, C. B e OSBORNE, M, A. The Future of Employment: How Susceptible Are Jobs to Computerisation? In: *Technological Forecasting and Social Change*. Vol. 114, 2017, p.254-280.
FGV – FUNDAÇÃO GETÚLIO VARGAS. Uso da Inteligência Artificial elevará Desemprego no País, Maio 2019 <eesp.fgv.br/noticia/uso-de-inteligencia-artificial-elevara-desemprego-no-pais > Acessado em 17 de Janeiro de 2020.
GIL, M. A. Como a inteligência artificial vai sair do controle e dominar o mundo. *Época Negócios*, Maio 2019, <epocanegocios.globo.com/Tecnologia/noticia/ 2019/05/como-inteligencia-artificial-vai-sair-do-controle-e-dominar-o-mundo. html>, Acessado em 1 Fevereiro de 2020.
IBM, Summit Supercomputer, 2018, <ibm.com/thought-leadership/summit-supercomputer> Acessado em 17 de Janeiro de 2020.
IEEE, IEEE Standard for Floating-Point Arithmetic, In: IEEE Std 754-2008, 2008.
LONG, L. N. and COTNER, C. F. A Review and Proposed Framework for Artificial General Intelligence. In: *2019 IEEE Aerospace Conference*, Big Sky, MT, USA, 2019, p. 1-10.
MANGAN, D. Lawyers could be the next profession to be replaced by computers, CBNC, 27 de Novembro de 2018 <cnbc.com/2017/02/17/lawyers-could-be-replaced-by-artificial-intelligence. html> Acessado em 17 de Janeiro de 2020.
MARSLAND, S. *Machine Learning* – An Algorithmic Perspective. Boca Raton: CRC Press, 2009.
MIAILHE, N. e HODES, C. The Third Age of Artificial Intelligence. In: *Field Actions Science Reports.* Vol 17, 2017, p. 6-11.
OLMOS, R. and TABIK, S. and HERRERA, F. Automatic handgun detection alarm in videos using deep learning. In: *Neurocomputing*. Vol. 275, 2019, p. 66-72.
OPEN PHILANTROPY. Potential Risks from Advanced Artificial Intelligence, Agosto 2015, < openphilanthropy.org/research/cause-reports/ai-risk> Acessado em 17 de Janeiro de 2020.
RUPP, K. CPU, GPU and MIC Hardware Characteristics over Time. 18 de Agosto de 2018. <.karlrupp.net/2013/06/cpu-gpu-and-mic-hardware-characteristics-over-time> Acessado em 17 de Janeiro de 2020.
RUSSEL, S. e NORVIG, P. *Inteligência artificial.* Trad. Regina Célia Simille de Macedo. 3ed. São Paulo: Elsevier, 2013.

SABBATINI, M. Novos modelos da percepção pública da ciência e da tecnologia: do modelo contextual de comunicação científica aos processos de participação social. In: *Intercom - IV Congresso Brasileiro de Ciência da Comunicação*, Campo Grande, 2004, p. 1-15.
SEARLE, J. R. *A Redescoberta da mente*. Trad. Eduardo Pereira e Ferreira. 2ed. São Paulo: Martins Fontes, 2006.
WALSH, F. *AI "outperforms" doctors diagnosing breast câncer*. BBC News, 2 de Janeiro de 2020, < bbc.com/news/health-50857759> Acessado em 17 de Janeiro de 2020.
YING, M. Quantum Computation, Quantum Theory and AI. In: *Artificial Intelligence*. Vol 174(2), 2010, p.162-176.

Submetido: 6 de fevereiro de 2020

Aceito: 4 de março de 2020