

Universidade Federal de São Carlos
Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

**Inteligência Artificial e teoria ator-rede em diálogo:
analisando atores e associações**

José Victor Rodrigues Catalano

São Carlos - SP
2023

JOSÉ VICTOR RODRIGUES CATALANO

Inteligência Artificial e teoria ator-rede em diálogo: analisando atores e associações

Dissertação apresentada ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, do Centro de Educação e Ciências Humanas, da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência, Tecnologia e Sociedade.

Orientador: Prof. Dr. Thales Haddad de Novaes Andrade

Coorientador: Prof. Dr. Bruno Rossi Lorenzi

São Carlos - SP
2023

Rodrigues Catalano, José Victor

Inteligência Artificial e teoria ator-rede em diálogo:
analisando atores e associações / José Victor Rodrigues
Catalano -- 2023.

138f.

Dissertação (Mestrado) - Universidade Federal de São
Carlos, campus São Carlos, São Carlos

Orientador (a): Thales Haddad Novaes de Andrade

Banca Examinadora: André Tosi Furtado, Sylvia

Iasulaitis

Bibliografia

1. Inteligência artificial. 2. CTS. 3. Teoria Ator-Rede. I.
Rodrigues Catalano, José Victor. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato José Victor Rodrigues Catalano, realizada em 07/11/2023.

Comissão Julgadora:

Prof. Dr. Thales Haddad Novaes de Andrade (UFSCar)

Prof. Dr. André Tosi Furtado (UNICAMP)

Profa. Dra. Sylvia Iasulaitis (UFSCar)

AGRADECIMENTOS

Expresso os meus sinceros agradecimentos a todos os atores que, de maneira direta ou indireta, contribuíram para esta pesquisa. Em primeiro lugar, gostaria de agradecer ao Prof. Dr. Thales Haddad Novaes de Andrade, meu orientador, pela disponibilidade, pelo apoio e, sobretudo, pela liberdade concedida para a realização deste estudo. Ao Prof. Dr. Bruno Rossi Lorenzi, meu coorientador, agradeço pelo apoio adicional e pelo acompanhamento atento ao longo do processo de escrita. À banca examinadora, tanto da qualificação quanto da defesa, composta pelo Prof. Dr. André Tosi Furtado e pela Profa. Dra. Sylvia lasulaitis, agradeço pelas valiosas sugestões que enriqueceram consideravelmente este trabalho. Estendo os meus agradecimentos ao Prof. Dr. Jadilson Astorino Marçola, pela disponibilidade de também participar de ambos os processos avaliativos. Agradeço também ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS), assim como a seus funcionários e coordenação, pelo suporte e pelas oportunidades proporcionadas. Agradeço aos docentes do PPGCTS pelas disciplinas oferecidas. À representação discente, agradeço pela colaboração na resolução das demandas acadêmicas. Gostaria de agradecer aos meus colegas Gabriel, Gabriella, Isabella e Léo pelo companheirismo demonstrado ao longo dessa jornada acadêmica, bem como à Vanessa, pelos conselhos e apoio nas questões burocráticas. Agradeço também aos meus pais, Aldomiro e Tânia, cujo apoio e incentivo foram fundamentais para o meu percurso acadêmico. Por fim, agradeço aos autores e pesquisadores cujas obras desempenharam um papel crucial no desenvolvimento deste estudo. Mais uma vez, muito obrigado a todos.

*"I often tell my students not to be misled by the name 'artificial intelligence' — there is nothing artificial about it. A.I. is made by humans, intended to behave by humans and, ultimately, to impact humans lives and human society."*¹

(Fei-Fei Li)

¹ "Costumo dizer aos meus alunos que não se deixem enganar pelo nome "inteligência artificial" - não há nada de artificial nela. A I.A. é feita por humanos, destinada a comportar-se como humanos e, em última análise, tem impacto na vida dos humanos e na sociedade humana." (tradução nossa)

RESUMO

O principal objetivo desta pesquisa consiste em aprofundar o entendimento crítico da Inteligência Artificial (IA) no contexto dos Estudos em Ciência, Tecnologia e Sociedade (CTS) por meio de uma abordagem científica. Buscamos aplicar a Teoria Ator-Rede (TAR) para analisar a complexa interação da IA com as dimensões científicas, tecnológicas e sociais. O propósito subjacente é reconhecer a IA como uma construção social, através da análise de sua dinâmica evolutiva e da identificação dos principais atores envolvidos nesse cenário, bem como de suas respectivas associações. De maneira complementar, apresentamos uma síntese cronológica que aborda os eventos mais significativos que moldaram a história da IA, incluindo a investigação das origens e motivações que impulsionaram seu desenvolvimento ao longo do tempo. Isso proporciona uma compreensão das influências históricas que contribuíram para a configuração da IA moderna. Por fim, concentramos nossos esforços em examinar a intrincada interação da IA com as dimensões do campo CTS, o que envolve uma investigação dos impactos éticos, políticos, econômicos e culturais da IA na sociedade. Reconhecemos que a influência da IA transcende o âmbito técnico e científico, estendendo-se a áreas cruciais da sociedade. Sob a perspectiva da TAR, destacamos a importância de compreender a IA não como uma tecnologia neutra, mas sim como um participante ativo em um cenário sociotécnico complexo e interconectado, influenciando e sendo diretamente influenciada por uma ampla gama de atores, sejam eles humanos ou não humanos.

Palavras-chave: Inteligência artificial; CTS; Teoria Ator-Rede

ABSTRACT

The main focus of this research is to deepen the critical understanding of Artificial Intelligence (AI) in the context of Studies in Science, Technology and Society (STS) through a scientific approach. We seek to apply Actor-Network Theory (ANT) to analyze the complex interaction of AI with scientific, technological and social dimensions. The underlying purpose is to consider AI as a social construction, through the analysis of its evolutionary dynamics and the identification of the main actors involved in this scenario, as well as their respective associations. In a complementary way, we present a chronological synthesis that addresses the most significant events that shaped the history of AI, including the investigation of the origins and motivations that contributed to its development over time. This provides an understanding of the historical influences that develop into the shape of modern AI. Finally, we focus our efforts on examining the intricate interplay of AI with dimensions of the STS field, which involves an investigation of ethical, political, economic, and cultural impacts of AI on society. We recognize that the influence of AI transcends the technical and scientific scope, extending to crucial areas of society. From an ANT perspective, we highlight the importance of understanding AI not as a neutral technology, but rather as an active participant in a complex and interconnected socio-technical scenario, influencing and being directly influenced by a wide range of actors, whether human or not humans.

Keywords: Artificial intelligence; STS; Actor-Network Theory

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 JUSTIFICATIVA.....	14
1.2 PERGUNTA E OBJETIVOS.....	17
1.2.1 Objetivo geral.....	17
1.2.2 Objetivos específicos.....	17
1.3 METODOLOGIA.....	18
2 CIÊNCIA, TECNOLOGIA E SOCIEDADE.....	20
2.1 DESCREVENDO O FAZER QUE FAZ FAZER: TEORIA ATOR-REDE.....	22
2.2 PARA ALÉM DA TÉCNICA: O PAPEL DOS OBJETOS NAS MEDIAÇÕES.....	32
3 INTELIGÊNCIA ARTIFICIAL.....	38
3.1 INTELIGÊNCIA ARTIFICIAL MODERNA.....	46
3.1.1 Aprendizado de máquina.....	48
3.1.2 Aprendizado profundo.....	51
4 A EVOLUÇÃO DA INTELIGÊNCIA ARTIFICIAL.....	58
4.1 PRIMEIRA ONDA DA IA: A ERA DE OURO (1956–1974).....	61
4.1.1 Primeiro inverno da IA: o descrédito da IA (1974–1980).....	65
4.1.2 Primeira primavera da IA: os sistemas especialistas (1980–1987).....	69
4.1.3 Segundo inverno da IA: a queda dos sistemas especialistas (1987–1993).....	72
4.2 SEGUNDA ONDA DA IA: REDES NEURAIAS (1993–2011).....	75
4.2.1 Segunda primavera da IA: o big data e aprendizado profundo (2011–2022)....	79
4.2.2 Verão da IA: a IA generativa (2022–presente).....	90
4.2.3 Perspectivas futuras.....	94
5 INTELIGÊNCIA ARTIFICIAL E TEORIA ATOR-REDE EM DIÁLOGO.....	96
5.1 MEDIANDO A CONTROVÉRSIA DO FUTURO DO TRABALHO.....	111
6 CONCLUSÕES.....	123
REFERÊNCIAS.....	129

1 INTRODUÇÃO

Atualmente, estamos testemunhando um esforço significativo para repensar e definir a era em que vivemos. Diversas abordagens estão sendo exploradas para compreender as transformações em curso na estrutura socioeconômica da sociedade contemporânea, e muitas delas convergem para a percepção de que a tecnologia digital desempenha um papel central nesse processo. De acordo com Silva Neto, Bonacelli e Pacheco (2020), é importante ressaltar que essas mudanças não estão sendo impulsionadas por uma única tecnologia isolada, mas sim por um conjunto interativo de tecnologias em constante evolução.

Os autores destacam que nos últimos dez anos, o avanço das tecnologias digitais tem se revelado como uma força motriz que tem gerado uma série de análises e debates relacionados a uma nova era dentro do contexto do capitalismo. Nesse sentido, eles introduzem o conceito de "sistema tecnológico digital", que abrange o período de 2010 em diante. Esse sistema é caracterizado pela interseção de três pilares fundamentais: computação em nuvem, *big data* e inteligência artificial (IA) (SILVA NETO; BONACELLI; PACHECO, 2020).

A interconexão desses três elementos - computação em nuvem, *big data* e IA - tem promovido uma metamorfose significativa em nossa sociedade. Essas tecnologias não apenas coexistem, mas também se complementam, criando um ambiente propício para mudanças radicais. A computação em nuvem fornece uma infraestrutura escalável e acessível, permitindo o armazenamento e o processamento massivo de dados, que é o cerne do *big data*. A IA, por sua vez, utiliza esses dados para aprender, tomar decisões e automatizar tarefas complexas, influenciando diretamente uma profunda transformação na forma como vivemos e trabalhamos (SILVA NETO; BONACELLI; PACHECO, 2020).

Conforme destacado por Silva Neto (2022), a IA finalmente encontrou aplicação em diversos setores após um período de antecipação significativo. Essa tecnologia vem sendo implementada com sucesso em áreas que abrangem a economia, a ciência, a cultura e muitos outros domínios. Essa implementação da IA marca o início de uma evolução contínua de um sistema tecnológico digital, cujas inovações estão desencadeando mudanças de grande magnitude tanto nos setores

tradicionais quanto nos emergentes. Como resultado, diversos campos de atuação estão sendo profundamente afetados por essas transformações em curso no cenário digital, que se mantém em constante mutação.

A IA, é um campo interdisciplinar que combina conhecimentos em ciência da computação, matemática, engenharia e ciências cognitivas (RUSSEL; NORVIG, 2013), com o objetivo de criar sistemas capazes de realizar tarefas que, até recentemente, eram exclusivas da inteligência humana (BRYNJOLFSSON; MCAFEE, 2014). Estas tarefas incluem o reconhecimento de padrões, a tomada de decisões, a compreensão de linguagem natural, a visão computacional, entre outras (RUSSEL; NORVIG, 2013). A IA está presente em uma variedade de aplicações cotidianas, desde assistentes virtuais em smartphones até sistemas de diagnóstico médico avançado, passando por carros autônomos e análise de dados em larga escala (TAULLI, 2019) (NEWQUIST, 2020).

No entanto, o rápido avanço da IA não está isento de desafios e dilemas éticos e sociais. À medida que as máquinas se tornam mais inteligentes e autônomas, questões sobre privacidade (ZUBOFF, 2019), segurança (SCHNEIER, 2016), viés algorítmico (BUOLAMWINI; GEBRU, 2018) e o impacto no emprego humano (BRYNJOLFSSON; MCAFEE, 2014) se tornam cada vez mais prementes. Esta dissertação busca também abordar essas questões cruciais sob a ótica da Teoria Ator-Rede, contribuindo para um melhor entendimento dos aspectos éticos e sociais da IA.

Os estudos em Ciência, Tecnologia e Sociedade (CTS), neste contexto, enriquecem a compreensão crítica da revolução tecnológica centrada na IA e sua interação com a sociedade contemporânea. A interseção entre a ciência, a tecnologia e a sociedade, destacando como esses três elementos estão intrinsecamente interligados e se moldam mutuamente um ao outro. Os estudos CTS, buscam analisar essa influência mútua, incluindo aspectos éticos, culturais, políticos e econômicos (SISMONDO, 2010).

A revolução tecnológica impulsionada pela IA não é apenas um desenvolvimento técnico; é também um fenômeno intrinsecamente entrelaçado com os tecidos sociais e culturais da nossa sociedade contemporânea (ZUBOFF, 2019).

A compreensão desse fenômeno complexo exige uma abordagem multifacetada, e os estudos CTS oferecem uma lente poderosa para examinar as implicações dessa revolução em todas as suas dimensões.

Em primeiro lugar, os estudos CTS nos permitem explorar como a IA não é apenas uma inovação tecnológica, mas também uma expressão das dinâmicas sociais e políticas que a impulsionam. A maneira como a IA é desenvolvida, implementada e regulamentada está intrinsecamente ligada às decisões e interesses das partes envolvidas, sejam empresas, governos ou comunidades. Essa perspectiva nos ajuda a entender como as tecnologias são moldadas por relações de poder, valores e agendas políticas, e como isso afeta sua adoção e impacto na sociedade (SISMONDO, 2008).

Além disso, os estudos CTS nos convidam a examinar as implicações éticas da IA. À medida que a IA se torna mais onipresente em nossas vidas, surgem questões cruciais sobre a responsabilidade social. Esses estudos nos incentivam a considerar como as escolhas tecnológicas podem perpetuar ou mitigar desigualdades e injustiças, e a desenvolver estratégias para uma adoção mais ética e responsável da IA.

Outro aspecto é o impacto da IA na economia e no trabalho. À medida que a automação impulsionada pela IA se expande, há preocupações sobre o deslocamento de empregos, a requalificação da força de trabalho e as consequências para a distribuição de renda (MCKINSEY; 2017, 2021). Uma análise completa dessas questões requer uma compreensão das interações complexas entre tecnologia, mercado de trabalho e políticas públicas, todas as quais são áreas de estudo dentro dos estudos CTS (SISMONDO, 2008).

Para essa análise interdisciplinar, na obra "Reagregando o Social: Uma Introdução à Teoria do Ator-Rede", Latour (2012) introduz o conceito de reagregar o social, uma expressão que busca reconfigurar e reavaliar a concepção de "social" na sociologia contemporânea. A proposta consiste em uma crítica à tradicional dicotomia entre o "social" e o "não social" na teoria sociológica, na qual o autor argumenta que tal dicotomia estabelece uma divisão arbitrária entre a humanidade e o mundo natural.

Ao adotar a perspectiva de reagregar o social, Latour (2012) busca ultrapassar a visão exclusivamente humana do mundo social e reconhecer a complexidade das redes de relações nas quais estamos inseridos. Ele propõe uma abordagem baseada em evidências e orientada para a prática, a fim de compreender essas redes, enfatizando a importância de estudar as interações entre humanos e não humanos como elementos integrantes da análise sociológica. Para o autor, devemos estender a noção de social para além das interações humanas e reconhecer a influência mútua entre esses atores na construção do mundo social.

A Teoria Ator-Rede, desenvolvida por Bruno Latour, Michel Callon e outros estudiosos, oferece uma perspectiva única para entender o papel da tecnologia e das redes sociais na sociedade (SISMONDO, 2008). Ela tem sido aplicada de maneiras diversas na interdisciplinaridade (até fora do CTS) para analisar uma ampla gama de fenômenos, desde a pesquisa científica e a inovação tecnológica até a política, a cultura e o meio ambiente (SISMONDO, 2008).

A TAR considera todos os elementos de uma rede, sejam humanos ou não humanos, como atores que desempenham papéis igualmente importantes na formação de ações e decisões (LATOURE, 2012). É uma perspectiva que busca entender como a sociedade é construída por meio das interações entre esses atores, atribuindo importância igual tanto aos seres humanos quanto aos objetos técnicos, instituições e outros elementos que participam dessas redes de relações (AKRICH, 2014). Neste contexto, a TAR nos convida a examinar a IA não apenas como uma criação humana, mas como um ator na rede sociotécnica que influencia e é influenciado por uma série de outros atores, incluindo seres humanos, algoritmos, instituições e dispositivos.

Ao considerar a visão de Latour (2012), busca-se analisar a IA como parte das dinâmicas sociais, reconhecendo seu papel e impacto nas interações que constituem o âmbito social. Em vez de separar a IA como algo “não social” ou externo às relações sociais, propõe-se analisar essa tecnologia ao entendimento do social. Ao analisar a IA a partir da TAR, busca-se compreender a IA como um ator relevante dentro das redes sociotécnicas. Portanto, é importante reconhecer que a IA desempenha um papel ativo e exerce influência nas dinâmicas sociais.

De acordo com a TAR, o objeto técnico não é neutro, mas sim modelado e influenciado pelas interações que mantém com outros atores na rede (AKRICH, 1997). Ele é considerado uma força que molda e é moldada pelas relações sociais e culturais em que está inserido (AKRICH, 2014). Além disso, a TAR enfatiza a importância da materialidade na construção da rede sociotécnica. Isso significa que os objetos técnicos influenciam o ambiente físico em que estão inseridos, assim como são influenciados pelos atores humanos e não humanos com os quais interagem (LATOUR, 2012).

Ao longo dos próximos capítulos, exploraremos os fundamentos históricos e teóricos da IA, seus principais métodos e técnicas, bem como sua aplicação em diferentes domínios. Através da lente da Teoria Ator-Rede, analisaremos também os desafios que a IA enfrenta e como a sociedade pode lidar com eles de maneira responsável e ética, considerando os atores e as redes envolvidos. Por fim, esta dissertação se propõe a contribuir para o conhecimento existente, oferecendo percepções críticas sobre o estado atual da IA e suas implicações para o futuro, sob a perspectiva da Teoria Ator-Rede.

1.1 JUSTIFICATIVA

A IA não se limita mais a tarefas simples, como reconhecimento de padrões ou respostas automáticas a perguntas. Ela evoluiu para sistemas altamente complexos e autônomos que podem aprender e se adaptar continuamente, tomando decisões sofisticadas com base em enormes conjuntos de dados e algoritmos avançados. A IA agora é capaz de entender o contexto, interpretar emoções humanas e até mesmo criar arte, música e literatura.

No entanto, à medida que a IA se infiltra em nossas vidas cotidianas, a falta de transparência e compreensão de seus processos internos assume proporções alarmantes. Muitas das escolhas significativas que têm impacto em nossas vidas, como concessões de crédito, contratações de emprego, apólices de seguro e outras, são determinadas por meio de algoritmos de IA e sistemas automatizados que permanecem inacessíveis ao conhecimento do público em geral. Esses algoritmos

são frequentemente considerados como sistemas complexos e impenetráveis, conhecidos como caixas pretas (PASQUALE, 2015).

Para a maioria das pessoas, a complexidade da IA a torna incompreensível, sendo frequentemente percebida como uma "caixa preta" onde as entradas e saídas são visíveis, mas o processo interno permanece obscuro (ZUBOFF, 2019). Essa metáfora de "caixa preta", conforme Latour (2000), sugere algo misterioso e inacessível, o que dificulta a transparência necessária para compreender como a IA toma decisões, quais dados utiliza e como pode perpetuar preconceitos ou viés indesejados. Quando os algoritmos de IA operam dessa forma opaca, seus resultados podem ser aceitos sem questionamento, mesmo quando resultam em discriminação, injustiça ou outros problemas éticos.

Latour (2000) argumenta que tratar algo como uma caixa preta leva as pessoas a aceitar seu funcionamento sem questionar ou investigar mais a fundo. No entanto, essa abordagem representa um risco significativo quando se trata de tecnologias tão poderosas como a IA, que podem influenciar profundamente a sociedade. Pois, enquanto a IA for encarada como uma caixa preta, perdemos a oportunidade de compreender seu impacto e de tomar decisões informadas sobre seu uso e desenvolvimento.

Para Latour (2000), a abertura das caixas pretas é crucial para compreender como a sociedade funciona e como as relações de poder são estabelecidas. Ele propõe desvendar o interior dessas caixas pretas, revelando os atores e os processos que as compõem. Ao fazer isso, é possível entender melhor como os interesses, as ideologias e as relações de poder moldam e influenciam essas entidades aparentemente autônomas.

Hoje, nos deparamos com um desafio de proporções significativas relacionado à maneira como rotulamos indiscriminadamente uma vasta gama de avanços tecnológicos sob o manto da "IA". Essa tendência é consequência do uso generalizado e muitas vezes superficial do termo, que agrupa diversas tecnologias e aplicações sob uma única etiqueta conceitual. Contudo, essa prática não apenas obscurece o verdadeiro significado da IA, mas também, de forma mais preocupante,

cria uma visão distorcida e simplista das implicações sociais e éticas dessa tecnologia.

O termo "inteligência artificial" é frequentemente empregado como um invólucro conveniente que abrange uma ampla gama de inovações tecnológicas, que variam desde algoritmos de aprendizado de máquina simples até sistemas complexos de processamento de linguagem natural e redes neurais profundas. Embora alguns desses sistemas possam exibir capacidades que se assemelham à inteligência humana, outros são altamente automatizados, realizando tarefas específicas sem compreensão ou raciocínio genuíno (MITCHELL, 2019). A generalização excessiva dessas tecnologias sob o rótulo "IA" pode criar a ilusão de que todas compartilham características e implicações similares, mantendo assim essa "caixa preta" intacta.

Essa generalização não crítica se torna particularmente inquietante quando confrontada com duas correntes de pensamento relacionadas à compreensão e influência da tecnologia na sociedade: o determinismo tecnológico e o positivismo tecnológico. Ambas essas abordagens tendem a simplificar em demasia a complexidade das interações entre tecnologia e sociedade, reduzindo o progresso tecnológico a uma visão unilateral e simplista, muitas vezes enclausurada em uma caixa preta.

Portanto, torna-se essencial adotarmos uma abordagem mais reflexiva e analítica em relação à IA, na qual consideremos sua trajetória de evolução, suas características intrínsecas e suas complexidades. A justificativa para isso é que as contribuições provenientes desta pesquisa visam agregar conhecimento tecnocientífico² no campo da IA, alinhando-se com o campo CTS, com o intuito de fortalecer e solidificar ambas as áreas.

² Latour (2000, p. 103), utiliza o termo "tecnociências" como uma maneira de enfatizar sua abordagem empírica e sociológica. O autor argumenta que a ciência está intrinsecamente conectada à sociedade e à política, negando a existência de uma ciência completamente independente e imune a influências políticas. E o êxito das tecnociências é resultado da organização política e da vontade humana, não de um reconhecimento universal de um conhecimento objetivo e incontestável.

1.2 PERGUNTA E OBJETIVOS

A questão central deste estudo é: "Como aprofundar a compreensão da evolução da inteligência artificial (IA) e de sua complexa interação com as dimensões científicas, tecnológicas e sociais?". Neste contexto, torna-se fundamental explorar não apenas o avanço tecnológico da IA, mas também o seu impacto nas esferas científicas e sociais, considerando as implicações éticas e os desafios emergentes que surgem dessa interseção dinâmica.

Para responder esta questão, começaremos traçando uma representação do estado atual da IA como um campo de estudo e uma tecnologia em constante evolução. No segundo estágio, descreveremos a evolução da IA por meio de um mapeamento cronológico, destacando os marcos significativos que moldaram a narrativa histórica deste campo.

Em seguida, concentraremos nossa atenção na análise da intrincada interação da IA com as dimensões do campo CTS. Esta investigação, sob a ótica da TAR, buscará analisar as várias maneiras pelas quais a IA foi concebida e aplicada ao longo do tempo, examinando seus agenciamentos, atores e associações correspondentes.

Por fim, encerraremos com uma análise das redes de controvérsias que envolvem a IA moderna, e examinaremos suas implicações prospectivas, levando em consideração os aspectos éticos, políticos, econômicos e culturais que permeiam a rede sociotécnica da IA. Isso nos permitirá entender o impacto profundo que a IA moderna exerce sobre o âmbito social por meio de suas controvérsias.

1.2.1 Objetivo geral

- Aprofundar a compreensão crítica da IA no contexto dos Estudos CTS.

1.2.2 Objetivos específicos

1. Mapear os antecedentes históricos que conduziram à evolução da IA;

2. Investigar o histórico do desenvolvimento da IA, considerando seus aspectos científicos, tecnológicos e sociais, bem como seus diversos atores e associações;
3. Analisar o impacto da IA moderna no contexto social, considerando, sob a ótica da TAR, suas principais controvérsias.

1.3 METODOLOGIA

Para aprofundar a compreensão da relação entre o desenvolvimento da IA e os Estudos CTS (Ciência, Tecnologia e Sociedade), o estudo proposto adotará uma abordagem de pesquisa do tipo indutiva, de natureza básica, com abordagem qualitativa, objetivos descritivos e técnicas bibliográficas. Esta metodologia permitirá uma investigação abrangente e sistemática da interação entre a IA e as dinâmicas socioculturais e políticas em que está inserida.

A pesquisa proposta adotará o método indutivo, que envolve a formulação de teorias e hipóteses com base em observações e análises de dados obtidos (MARCONI; LAKATOS, 2021). Seguirá uma abordagem qualitativa, concentrando-se na interpretação do fenômeno do desenvolvimento da IA com objetivo descritivo. Buscaremos descrever e compreender esse fenômeno por meio da aplicação da Teoria Ator-Rede, a fim de examinar como a IA influencia e é influenciada por fatores sociais, econômicos e políticos.

Já o levantamento bibliográfico será dividido em quatro seções distintas. Na primeira seção, conduziremos uma revisão bibliográfica sistemática explorando os conceitos essenciais no campo CTS. Esse estágio estabelecerá uma base sólida para nossa investigação sobre a interdependência entre ciência, tecnologia e sociedade.

Na segunda seção deste estudo, será conduzida uma investigação sistemática do desenvolvimento teórico do campo CTS, com ênfase nas principais contribuições de Bruno Latour e Madeleine Akrich no contexto da Teoria Ator-Rede. Esta análise se concentrará na compreensão da mediação de objetos técnicos com o objetivo de examinar a maneira como esses autores abordam a interação entre

atores humanos e não humanos interagem e se relacionam em redes complexas para criar a realidade social, estabelecendo assim uma sólida base teórica e metodológica para orientar a pesquisa.

A terceira seção deste estudo se dedica à descrição e caracterização da IA e da IA moderna. Nesse contexto, procedemos à exploração dos conceitos e características mais recentes no âmbito da IA como campo de pesquisa em constante evolução. Além disso, procuramos caracterizar a IA moderna por meio da análise de suas principais abordagens. A fim de fundamentar nossa análise, recorreremos a uma seleção sistemática de fontes, incluindo literatura especializada em forma de livros de referência, teses e dissertações provenientes de programas de pós-graduação, bem como artigos criteriosamente selecionados em repositórios acadêmicos de destaque, tais como o Portal de Periódicos da CAPES, *Scientific Electronic Library Online (SciELO)*, *Scopus*, *Web of Science*, *ScienceDirect*, *IEEE Xplore*, *ACM Digital Library*, *SSRN* e Google Acadêmico.

Para a execução do levantamento bibliográfico da produção científica, implementamos uma estratégia de busca detalhada, utilizando os seguintes termos de pesquisa: "inteligência artificial" ou "*artificial intelligence*", "aprendizado de máquina" ou "*machine learning*" e "aprendizado profundo" ou "*deep learning*". Esses termos orientaram nossa pesquisa de maneira a abranger amplamente o escopo da IA moderna, proporcionando uma base sólida para nossas análises e conclusões.

Na quarta seção desta pesquisa, voltaremos nossa atenção à análise histórica do desenvolvimento da IA, traçando seu progresso ao longo do tempo com base nas contribuições de pesquisadores renomados, incluindo McCorduck, Mitchell, Newquist, Norvig, Russel e Taulli. Esta investigação será conduzida por meio de uma revisão abrangente da literatura que aborda o desenvolvimento da IA no período de 1956 a 2020, com o propósito de identificar a trajetória dessa tecnologia ao longo dos anos.

Além disso, para enriquecer nossa análise com informações atualizadas, empregaremos as mesmas fontes de dados previamente mencionadas para a seleção de artigos recentes que abordam a palavra-chave "inteligência artificial generativa" ou "*generative artificial intelligence*". Essa estratégia nos permitirá

examinar o período mais recente, de 2020 a 2023, e compreender como a IA generativa se desenvolveu nesse intervalo temporal.

Por fim, procederemos à análise do desenvolvimento da IA sob a perspectiva da TAR, com a descrição detalhada de seus agenciamentos, mediações e controvérsias. Essa abordagem nos permitirá evidenciar que a IA transcende a sua mera condição de instrumento técnico, assumindo o papel de ator na sociedade contemporânea, contribuindo tanto para a sua configuração quanto sendo influenciada por diversos atores.

Para realizar a presente análise, procederemos à identificação dos marcos críticos no progresso da IA já levantados anteriormente durante a análise bibliográfica. Adicionalmente, examinaremos as questões críticas suscitadas pela IA nos domínios da ciência, tecnologia e sociedade, fundamentados em uma revisão da literatura crítica contemporânea. Por fim, exploraremos a controvérsia concernente ao "futuro do trabalho" sob a ótica da TAR e das análises efetuadas por autores críticos ao tema.

Nesta fase do estudo, adotaremos a metodologia indutiva. Através deste método, erigiremos uma base argumentativa sólida que nos possibilitará estabelecer vínculos coerentes entre os resultados alcançados e as premissas iniciais subjacentes à nossa investigação. Este processo de raciocínio dedutivo culminará em conclusões firmemente embasadas (MARCONI; LAKATOS, 2021).

2 CIÊNCIA, TECNOLOGIA E SOCIEDADE

A expressão "ciência, tecnologia e sociedade" (CTS) denota um campo interdisciplinar de estudo que se dedica a analisar as intrincadas relações entre ciência, tecnologia e a sociedade humana, considerando tanto os elementos sociais que influenciam o desenvolvimento da ciência e da tecnologia quanto as ramificações sociais e ambientais resultantes dessas atividades. Os estudos CTS desempenham um papel fundamental na obtenção de uma compreensão abrangente do papel desempenhado pela ciência e tecnologia na sociedade contemporânea, bem como na formulação de decisões embasadas sobre sua

utilização e impacto (PALÁCIOS *et al.*, 2009). A pesquisa em CTS abarca uma vasta gama de tópicos essenciais para a compreensão das influências mútuas e dos efeitos da ciência e tecnologia no ambiente social em constante evolução (SISMONDO, 2008).

No campo dos CTS, um dos principais focos de investigação é a análise da relação entre ciência, tecnologia e sociedade, abordando tanto suas origens na sociedade quanto os impactos que geram nas esferas sociais e ambientais (PALÁCIOS *et al.*, 2009). Além disso, os estudos CTS se concentram na análise do desenvolvimento e adoção de tecnologias, examinando todo o ciclo de inovação tecnológica, desde sua concepção e criação até sua integração na sociedade. Nesse contexto, são avaliados os fatores que promovem ou limitam a aceitação de novas tecnologias (SISMONDO, 2008).

É importante ressaltar que o contexto social, político e econômico desempenha um papel crucial na configuração da ciência e da tecnologia. Portanto, os estudos CTS investigam detalhadamente como esses fatores influenciam as escolhas de pesquisa, as decisões relacionadas ao desenvolvimento tecnológico e as políticas relacionadas à ciência e à tecnologia (PALÁCIOS *et al.*, 2009).

Em vez de conceber a ciência e a tecnologia como processos independentes e estritamente racionais, os estudos CTS fomentam uma perspectiva renovada da atividade científica, que não se baseia em essencialismos e está profundamente enraizada no contexto social (PALÁCIOS *et al.*, 2009). Nesse contexto, Sismondo (2010) argumenta que os fatos científicos e os artefatos tecnológicos não são entidades objetivas e neutras, mas, ao contrário, são moldados por esses contextos sociais e culturais em que surgem. Isso envolve examinar como o conhecimento científico é influenciado por valores sociais, interesses e dinâmicas de poder na sociedade. Contestando a ideia de determinismo tecnológico em que a tecnologia é um motor autônomo da mudança social, independente da ação humana. Em vez disso, ele destaca a coevolução entre tecnologia e sociedade, enfatizando como ambas se influenciam mutuamente.

Essa intersecção entre CTS, conforme destacado por Sismondo (2008) e Palácios *et al.* (2009), oferece uma perspectiva enriquecedora para a análise das

intrincadas relações que permeiam o campo da IA e seu impacto no mundo que a cerca. Esta abordagem multidisciplinar não se limita a simplesmente admirar as realizações tecnológicas da IA, mas sim busca compreender suas consequências nas esferas sociais, éticas, políticas e econômicas. Ao adotar essa abordagem integrativa, estamos ampliando nossos horizontes e explorando os dilemas e desafios que acompanham o avanço constante da IA, enxergando-a como uma fusão entre o que é ciência, técnica e cultura.

Neste contexto, à medida que abrimos as portas para a colaboração e a intersecção entre uma variedade de campos do conhecimento, estamos pavimentando um caminho que transcende as barreiras tradicionais das disciplinas acadêmicas. Esse cruzamento de fronteiras intelectuais tem o poder de nutrir a inovação, estimular a colaboração e nos habilitar a enfrentar os desafios cada vez mais interligados do nosso tempo. Isso inclui a tarefa complexa de compreender e lidar com tecnologias em constante evolução, como é o caso da IA.

A IA, assim, não é apenas um artefato, mas também um agente de transformação social e cultural. Ao entendermos plenamente o impacto da IA sob a lente da CTS, podemos trabalhar de forma mais eficaz para garantir que seus benefícios sejam distribuídos de maneira equitativa, que os dilemas éticos sejam abordados de maneira responsável e que as políticas sejam moldadas para promover o bem-estar coletivo.

2.1 DESCRREVENDO O FAZER QUE FAZ FAZER: TEORIA ATOR-REDE

Entre o final dos anos 1970 e início dos anos 1980, na França, desenvolveu-se uma nova corrente de pensamento no campo da sociologia, a Sociologia Pragmática Francesa. Autores como Bruno Latour, Luc Boltanski, Laurent Thévenot, Daniel Cefaï e Francis Chateauraynaud emergiram nesse contexto e, por meio de suas abordagens, buscaram compreender a sociedade e as ações humanas de uma forma diferente das correntes estabelecidas da época. Para esses autores, "o social deixa de ser o elemento explicativo das coisas e torna-se aquilo que deve ser explicado a partir de relações e movimentos problemáticos" (CORRÊA, 2014, p.39). Essa nova perspectiva considera as ações de atores não humanos

como parte integrante e influente na sociedade, permitindo assim o estudo dos seres humanos e não humanos como produtores do social e da realidade resultante de suas ações (CORRÊA, 2014) (SALGADO, 2022).

Essa distinção de abordagem se baseia na compreensão da ação como um elemento plural, conforme defendido por Salgado (2022), ao relacionar a tradução do termo grego "*pragma*" com as palavras "ação" e "fazer". De acordo com Andrade (2000), no pensamento grego antigo, não havia uma distinção clara entre o agente (sujeito) e o receptor da ação (objeto), uma vez que o agir, o pensar e o ser estavam em harmonia (SALGADO, 2022, p.11).

Com foco atribuído à ação e aos diversos elementos envolvidos por ela, podemos compreender que essas, são realizadas por múltiplos e heterogêneos atores que envolvem um coletivo, híbrido, que não se limita somente aos seres humanos, e trata de forma igual os não humanos, defendendo que o social é um agrupamento dinâmico de ações, movimentos e deslocamentos de atores (SALGADO, 2022).

Corrêa (2014, p.39) afirma que a Sociologia Pragmática Francesa trata que o social é formado por meio de um "emaranhado de relações e associações dinâmicas" que são formadas através de "situações problemáticas". Neste caso, o social deixa de ser algo estabelecido e sim estabelecido através de um conjunto de situações que emergem através da ação dos envolvidos, assim, o social se torna o problema que emerge a partir das formas de resolução dessas situações (CORRÊA, 2014).

O social, nesse sentido, deixa de ser tratado como uma "coisa", ou seja, uma substância dotada de características e traços positivos a partir dos quais os elementos não sociais (como a religião, a arte, a ciência e a própria sociologia!) ganham sentido, e torna-se um conjunto de movimentos, associações, transformações, enfim, um fluxo tornado apreensível a partir de *affaires* (Boltanski e Claverie, 2007; Chateauraynaud, 2011), momentos críticos (Boltanski, 1990; Boltanski e Thévenot, 1991), situações problemáticas (Cefaï, 2006), momentos de prova e controvérsias (Chateauraynaud e Bessy, 1995; Latour, 2006 e 2010) (CORRÊA, 2014, p.39).

Neste contexto, temos autores como Bruno Latour, Michel Callon, John Law, Madeleine Akrich e Annemarie Mol, que por meio de seus estudos, desenvolveram uma abordagem metodológica a fim de descrever a fabricação dos fatos científicos utilizando a principal idéia da Sociologia Pragmática Francesa, a ação; formando um conjunto de estudos focados no ambiente e no contexto em que a ação está em curso. Tornando parte do âmbito dos Estudos em Ciência, Tecnologia e Sociedade (ECTS) e também um ramo das Sociologias Pragmáticas, denominado de Teoria Ator-Rede (SALGADO, 2022).

Alguns exemplos que consolidaram a TAR na época, estão nos trabalhos de Latour de 1979 e 1984; de etnografia de um laboratório de neuroendocrinologia no Instituto Salk na Califórnia e a posteriori um estudo sobre a pasteurização da França com foco na vacina antraz. Callon também desenvolveu um estudo sobre a experiência do desenvolvimento do veículo elétrico na década de 1970 na França e sobre a domesticação das vieiras na baía de St.-Brieuc de 1896. Já Law em 1986 por meio do estudo das navegações oceânicas portuguesas no século XVI desenvolveu toda a rede que formou em volta da expansão marítima portuguesa que possibilitou estabelecer uma rota comercial para a Índia (LATOUR, 1988) (LATOUR; WOOLGAR, 1997) (CALLON, 1980, 1986) (LAW, 1986).

A Teoria Ator-Rede (TAR) tem seu início no final da década de 1970 e início de 1980, seus principais precursores são os sociólogos franceses Michel Callon e Bruno Latour. A obra, "A vida de laboratório: a produção dos fatos científicos", escrita por Bruno Latour em parceria com o sociólogo Steve Woolgar em 1979, introduziu as primeiras ideias metodológicas. Artigos como "Desparafusando o grande Leviatã: como atores estruturam a realidade de maneira macro e como sociólogos os ajudam a fazer isso" de Callon e Latour de 1981 e "Por uma sociologia relativamente exata" de 1983, serviram como textos introdutórios para a metodologia. Mas somente em 1986 no artigo "A sociologia de um ator-rede: o caso do veículo elétrico" que Callon cunha este termo (SALGADO, 2022).

Apresentada como uma "abordagem metodológica para descrever a fabricação dos fatos científicos" (SALGADO, 2022, p.32) a TAR, foi desenvolvida no

campo dos Estudos da Ciência, Tecnologia e Sociedade (ECTS), proveniente de um grupo homônimo (*Sciences, technologies et société*) do Conservatório Nacional de Artes e Ofícios da França, datado de 1977, com o propósito de se diferenciar tanto dos estudos funcionalistas quanto dos estruturais que eram realizados pela Sociologia da Ciência na época (SALGADO, 2022).

Para entender a fabricação dos fatos científicos, Latour em suas obras "A vida de laboratório: a produção dos fatos científicos" de 1979 e "Ciência em ação: como seguir cientistas e engenheiros sociedade afora" de 1987. Nessas obras Latour relata de forma descritiva a prática cotidiana dos cientistas e engenheiros, com isso, ele defende que a ciência é uma ação e deve ser estudada no momento em que a mesma está ocorrendo, portanto, seu *locus* seria o cotidiano dos laboratórios (SALGADO, 2022).

Ao estudar a ciência em sua ação, Latour (2000) trata que a realidade da fabricação dos fatos científicos é de caráter social, definido por meio da interação entre o que ele define por actantes, que seriam atores humanos e não humanos que integram todo o contexto da atividade de laboratório (LATOUR, 2000). Os atores, desencadeiam uma "dinâmica de mútua afetação (ser levado por outros a agir e levar outros a agir)" (SALGADO, 2022, p. 33), entendendo a ação neste contexto da TAR como "um fazer que faz fazer" (SALGADO, 2022).

Sob a perspectiva da TAR, Latour (2000, 2012), define o actante como um ator, seja ele humano ou não humano, que desempenha uma ação na rede. Pois segundo o mesmo, a "ação não é uma propriedade dos humanos, mas de uma associação de actantes" (Latour, 1994, p. 35). Sendo assim, a ação transcende o social e o material, ao rejeitar tanto o antropocentrismo quanto o materialismo, demonstrando a equidade entre o homem/objeto ou objeto/homem. Nesta visão temos o ator/actante como "aquele" ou "aquilo" que executa uma ação, desempenhando uma mediação entre os demais atores em uma rede, sejam eles humanos ou não humanos, capaz de alterar as relações da rede (LATOUR, 2000, 2001, 2012) (SANTAELLA; CARDOSO, 2015).

O segredo é definir o ator com base naquilo que ele faz - seus desempenhos - no quadro dos testes de laboratório. Mais tarde, sua competência é deduzida e integrada a uma instituição. Uma vez que, em inglês, a palavra actor (ator) se limita a humanos, utilizamos muitas vezes "actant" (atuante), termo tomado a semiótica, para incluir não-humanos na definição (LATOURE, 2001, p. 346).

Mol (2010) complementa que um ator é fácil de definir, ele atua, faz algo que faz a diferença, e ao eliminar esse ator do cenário (que no caso seria a rede) os outros atores teriam dificuldade para substituir as ações decorrentes deste ator. Mol ainda aponta que o objetivo principal da TAR, não está em questionar aos atores a origem de suas ações, nem seus pontos de partida, mas sim, como essas ações se movimentam no campo de estudo das redes, pois os efeitos são essenciais (MOL, 2010).

A noção de rede é formada por meio das transformações geradas pelos conjuntos de interações entre os atores que de forma dinâmica e contínua, que modificam um ao outro. Latour (2012) define como “uma narrativa, descrição ou uma proposição na qual todos os atores fazem alguma coisa e não ficam apenas observando” (LATOURE, 2012, p. 189). E com isso, essas interações não servem simplesmente de meio de transporte sem efeitos de transformação, são ações, que podemos comparar a pontos de desvio, eventos ou até mesmo a origem de novas traduções. Sendo a descrição a melhor maneira de entender como lidar com as dinâmicas sociais que não podem ser tratadas por métodos expositivos.

Essas interações dinâmicas que ocorrem entre as ações dos atores são as ações que formam as redes e não um conjunto de padrões facilmente identificáveis. Latour (2012) compara as redes a um texto bem escrito que seja capaz de engajar cada ator a estimular outros a realizarem diversas ações. Elas são estabelecidas por um singular elemento constitutivo - os nós -, estes, gerados através das mediações de atores heterogêneos que ao se comunicarem, convergem seus interesses reformulando todo o processo e gerando redefinições, novos controles e ordenamento sociais (LATOURE, 2000, 2012). Latour (2012) defende a existência da

estrutura, sistema ou o contexto, mas a mesma é formada somente ao final dos processos de tradução entre os atores, sendo um resultado das mediações.

Latour (2012) define dois termos técnicos essenciais para entendermos essas ações, denominados de mediadores e intermediários, estes, fazem grande diferença ao serem adicionados aos meios de produzir o social. Ainda salienta que no início ambos podem ter o mesmo significado, mas no fim de suas ações, que o conduzirão a diferentes territórios.

Os mediadores, não podem ser relativizados como apenas um, pois possuem características antagônicas complexas ao se mover durante suas ações (LATOUR, 2012). Portanto, “os mediadores transformam, traduzem, distorcem e modificam o significado ou os elementos que supostamente veiculam” (LATOUR, 2012, p. 65). Assim, os mediadores, dependem do momento em que a ação está ocorrendo e também do seu ponto de vista. Sendo que um mediador, em algum momento de sua existência, já foi um intermediário, e vice versa. Nesse contexto, o mediador faz parte de um elemento importante da TAR, eles possibilitam as associações, pois, por meio delas conseguimos entender como as ações e relações dos mediadores que transformam as redes que os mesmos estão inseridos (LATOUR, 2012).

Um intermediário é aquilo que “transporta significado ou força sem transformá-los” (LATOUR, 2012, p. 65) ele é cristalizado ao definir o que entra e o que sai, sendo comparada à uma caixa preta que trabalha com unidade (LATOUR, 2012).

O conceito de caixa preta é uma expressão proveniente da cibernética que define que sempre quando há algo complexo demais, se desenha uma caixa preta a fim de omitir todo o processo. Esse conceito foi adaptado na TAR para definir que quando as mediações entre os atores, já finalizadas, formam uma caixa preta fechada, e todo o processo de construção do conhecimento é abandonado (LATOUR, 2000).

Assim, na medida que se afirma que não há necessidade de se preocupar com as discussões sobre determinada ciência e tecnologia e pular diretamente para seu uso, sem a preocupação de fato com a comprovação científica, suas mediações

e o processo de construção dos fatos, uma caixa preta está fechada (LATOUR, 2000).

Latour (1986) define a associação como um “poder” que deve ser tratado como uma consequência de uma ação e não algo que se possui. Essas associações, fazem parte central do que Latour (2012) defende como sociologia das associações, que é o conceito base para a TAR. O autor defende que devemos observar como são realizadas as associações, pois por meio da identificação das mesmas, são constituídas por grupos de atores em conexão que ao se interagirem, revelam as redes de mediadores que estruturam um sistema (LATOUR, 1986, 2000, 2012).

Com isso, além de identificar essas associações, conseguimos compreender a estrutura de uma rede. Latour (1986) ainda pontua que o social emerge a partir dessas associações, essas que são compostas por atores produzindo ações localizáveis, esses sendo humanos e não humanos. E o principal propósito da TAR está em buscar e analisar como se formam essas associações a fim de descrevermos o social (LATOUR, 1986).

Latour (2012) propõe o uso da TAR para a escrita de relatos, por meio de uma narrativa, descrição ou proposição em que os atores executam ações relevantes que transformam toda a rede. “[...] na qual todos os atores fazem alguma coisa e não ficam apenas observando.” (LATOUR, 2012, p.189). Nessa ação que transforma, cada ponto do texto pode se tornar uma encruzilhada, evento ou origem de uma nova mediação. Sendo assim tornando-se um mediador que por meio desse relato é que as redes são delimitadas. A rede nesse ponto se torna uma ferramenta para avaliar os movimentos e as especificidades dos próprios relatos a fim de auxiliar a descrever algo (LATOUR, 2012).

Latour (2012) defende que a formação de grupos e anti-grupos são resultado das ações dos próprios atores, esses humanos e não humanos (objetos), por isso, eles são capazes de definir todo o seu contexto social por meio do conjunto de associações que se formam entre eles. Diferente da sociologia tradicional, não podemos encarar os grupos como algo pré-estabelecido e contextualizado pelo pesquisador, pois somente por meio do rastreamento das conexões que ocorrerá a

formação dos grupos já que o dinamismo das mediações mudam o contexto das redes (LATOOUR, 2012). "Por razões científicas, políticas e mesmo morais, não convém que os pesquisadores definam antes dos atores, e no lugar deles, o elemento básico de que o mundo social é feito" (LATOOUR, 2012, p. 69).

No que tange a natureza das ações e sua descrição, a mesma não é consciente e sim articulada. Latour (2012) define que o "ator" da expressão "ator-rede" não é a fonte do ato e sim o alvo de um conjunto de outros atores que vão em sua direção, tornando-se a própria rede. Pois a ação é deslocada, isto é, está em movimento que leva o ator a agir. "A ação é tomada de empréstimo, distribuída, sugerida, influenciada, dominada, traída, traduzida." (LATOOUR, 2012, p. 76).

Latour (2012) pontua que uma ação é assumida por outras e que o social é formado por movimentos, deslocamentos, transformações, translações e registros; destaca-se o social como um recurso taquigráfico que descreve aquilo que já foi consolidado. Sendo o social uma associação momentânea entre atores que se aglutina e assume novas formas. Nisso, faz-se necessário definir que a ação social não é apenas assumida por atores humanos, mas também por objetos (não humanos), formando associações híbridas. Assim, os objetos surgem para explicar as assimetrias, expressar as relações de poder, simbolizar as hierarquias sociais, agravar ou desagravar as desigualdades sociais, transportar poderes, objetivar a igualdade e materializar as relações (LATOOUR, 2012).

Um conceito importante da TAR, é a mediação. Latour (1994), relaciona-se a uma nova ação que, no seu ato, modifica mutuamente os atores envolvidos. Ou mesmo um "deslocamento, deslize, invenção, mediação, a criação de uma conexão que não existia antes e que, em algum grau, modifica os dois elementos ou agentes" (Latour, 1994, p. 32). A mediação torna-se o produto da relação entre os atores. Santaella e Cardoso (2015) definem que a mediação é o que forma o conceito de hibridismo nas redes, ou "ator híbrido", pois, quando um actante, "faz o outro fazer", várias associações entre humanos e não humanos são formadas. (LATOOUR, 1994)

A mediação neste caso, está relacionada com o compartilhamento mútuo entre os atores que, no momento da ação, dividem suas responsabilidades à medida

que respeitam as ações dos demais atores envolvidos, formando assim uma composição. Pois, somente com a soma de todos os envolvidos nas ação se obtém um sentido na mediação (SANTAELLA; CARDOSO, 2015).

O conceito de mediação portanto, é tratado como um componente inexorável à TAR, pois permite a inclusão dos híbridos no contexto do social, já que funciona como um "fio condutor" para a teoria ator-rede ao envolver esses atores humanos e não humanos (CARDOSO; SANTAELLA, 2020, p.141). Essa ação conectiva entre os agentes mediadores é o que agencia o coletivo de atores em ação, pois as ações dos atores envolvidos no coletivo são simetricamente consideradas ao incluírem a relação híbrida entre humanos e não humanos no contexto do social que movimentam a própria rede (CARDOSO; SANTAELLA, 2020).

Atentar à simetria ou o tratamento simétrico dos atores é muito importante, pois somente quando igualamos o tratamento dos atores humanos e não humanos, podemos observar como essas associações se formam e se compõem. Goldman *apud* Cardoso e Santaella (2020, p. 144) recorda que a simetriação não tem relação em “colocar todos os atores (humanos e não humanos) em pé de igualdade, mas em evitar as dicotomias assimétricas e muitas vezes prematuras da modernidade”. Essa simetria é tratada inicialmente por Latour (1994) ao pontuar que a ação da técnica também cria condições para o ser humano. (LATOURE, 1994) (SANTAELLA; CARDOSO, 2015) (CARDOSO; SANTAELLA, 2020)

Ao analisarmos uma rede, se faz necessário também, observar a mesma em seu processo de produção dos fatos que possam ser considerados controversos, sendo assim, antes de fechar a caixa preta, e a presença da visibilidade dos fatos da rede que está sendo analisada. Em sua obra “Ciência em Ação”, ao analisar a ciência por meio de suas práticas e das práticas dos cientistas envolvidos, Latour (2000) desenvolve durante os capítulos, sete regras metodológicas que devem ser aplicadas quanto à observação da produção dos fatos científicos (LATOURE, 2000, p. 421-422):

1. O objeto de estudo não pode estar consolidado, devemos atentar ao momento de construção das caixas-pretas e seu fechamento, objetivando o mesmo em uma ação, atentando-se no objeto de controvérsia. Tendo como

base o ideal de estudar a ciência atual em seu processo de produção, em lugar das sancionadas (cristalizadas) (LATOUR, 2000)..

2. A natureza de um fato é efeito dos processos de circulação dos mesmos em uma rede, tendo a objetividade ou subjetividade como um efeito dessa interação. Assim, "não devemos procurar as qualidades intrínsecas de qualquer afirmação, mas sim todas as transformações" (LATOUR, 2000, p.99). E o ponto de partida está em observar o cenário em que a alegação (controvérsia) se encontra.
3. A estabilização da natureza do fato ocorre na resolução das controvérsias e não na causa de um fato controverso;
4. A estabilização da sociedade também ocorre na resolução das controvérsias;
5. Todos os nós que são formados durante os processos de mediação devem ser observados com o mesmo rigor e devem seguir a mesma simetria;
6. A rede que é construída pela situação de controvérsia e deve ter sua extensão aplicada ao que tange o objeto de acusação por seu ângulo e direção de deslocamento não se preocupando com irracionalidades, lógicas nem estruturas sociais e sim se aproximando da abordagem simétrica em sua observação;
7. Durante o rastreamento da rede é necessário analisar como ocorrem as inscrições e não atribuir um papel especial aos nós que possuem características de maior estabilidade.

Latour (2012) considera que ao estudarmos o processo de inovação os objetos ganham maior visibilidade, pois os mesmos estão em pleno desenvolvimento abrindo margem para maior visibilidade dos fatos nos ambientes, pois suas conexões geram maior exposição gerando assim as controvérsias. O autor pontua que ao visitarmos uma "obra em construção" podemos ter uma melhor visibilidade dos fatos e das respectivas relações entre os humanos e não humanos. Sendo necessário encontrar um equilíbrio que possibilite trabalhar com fatos de forma

rígida, mas também, se atentar a maneabilidade de fluxos dinâmicos no momento em observamos a produção dos fatos através dos movimentos de mediação (LATOURE, 2012).

Usualmente, a grande vantagem de visitar locais de construção é que eles oferecem um ponto de observação para se testemunhar a ligação entre seres humanos e seres não humanos. Assim que atolam o pé na lama, os visitantes são facilmente surpreendidos pelo espetáculo de todos os participantes trabalhando arduamente no momento de sua metamorfose mais radical (LATOURE, 2012, p. 131).

Mol (2010) defende assim a natureza flexível da TAR a um “repositório aberto e adaptável” (MOL, 2010, p. 265), formado por uma multiplicidade de dispositivos de coordenação. A autora argumenta que o ponto forte da TAR é a adaptabilidade, garantindo sustentabilidade e diversas aplicações em pesquisa. Ela, portanto, a define como 'uma tradição teórica selvagem e criativa'.

Porque a TAR consegue explorar o mundo de diferentes formas e vivenciá-lo em conjunto. São comparados referenciais teóricos de diversos autores como dispositivos de coordenação com preocupações e tópicos, incluindo o estudo de ações e transições que ocorrem de um contexto para outro. Esses dispositivos ajudam a analisar as ações de um ator na tradução do que está acontecendo durante a ação em rede. Dessa forma, o mesmo ator poderá observar como essas mudanças são realizadas, seus movimentos e suas intervenções na rede. De certo modo, essas ações se dividem em redes e as perspectivas são alteradas ou agregadas nesse movimento contínuo (MOL, 2010).

2.2 PARA ALÉM DA TÉCNICA: O PAPEL DOS OBJETOS NAS MEDIAÇÕES

Os objetos técnicos desempenham uma função significativa na interação entre seres humanos e seu ambiente. Isso se deve ao fato de que esses objetos estão imersos tanto no domínio físico quanto no contexto social, assumindo papéis ativos como atores nessa relação. É essa dinâmica de interação entre humanos e não humanos que possibilita que os objetos técnicos desempenhem um papel na

construção do mundo social e atuam como mediadores nas relações interpessoais, sendo, de fato, os intermediários essenciais em todas as nossas interações com a realidade circundante (AKRICH, 2014).

Vários autores se interessaram pelo modo como a técnica pode prolongar, no espaço e no tempo, a ação política. Nós propusemos inverter essa hipótese e mostrar como, longe de representar mais que um apêndice sobre um dispositivo político preexistente, os objetos técnicos possuem um conteúdo político no sentido de que eles constituem os elementos ativos de organização das relações dos homens entre eles e com seu ambiente (AKRICH, 2014, p. 161).

Segundo Akrich (2014, p. 162), os objetos técnicos formam um paradoxo pois, “embora claramente definido em seu aspecto físico, não é curiosamente menos imperceptível” pois, por mais que eles retornem ao seu fim para qual foi projetado, eles deixam um lastro intermediário na cadeia de associações entre humanos e não humanos. Esses objetos são, portanto, compostos por forças de várias naturezas, devido às quais uma série de relações de elementos heterogêneos da sociedade são conectadas e associadas entre si (AKRICH, 2014).

Um objeto técnico define não somente os atores e as relações entre esses atores, mas deve, para continuar a funcionar, estabilizá-los e canalizá-los: ele estabelece os sistemas de causalidade que se apoiam sobre os mecanismos de rarefação de sentidos (AKRICH, 2014, p. 179).

Akrich (2014) analisa que os objetos técnicos podem se tornar mediadores, atingindo outros atores e atribuindo-lhes papéis específicos em uma rede, definindo assim novas relações dentro do contexto em que está inserido. Esses objetos têm a capacidade de medir relacionamentos, estabilizar hierarquias, estabelecer normas e comportamentos em redes e operar a partir de associações entre atores técnicos e humanos. Formando uma relação técnica, econômica e social estável e tem como produto final desta mediação o que Akrich (2014) define de 'roteiro' ou 'cenário'. Evidenciando os “mecanismos elementares de ajustamento recíproco do objeto técnico e de seu ambiente” (AKRICH, 2014, p. 165), porque os objetos técnicos

transcendem simultaneamente os elementos técnicos e sociais durante a sua criação (AKRICH, 2014).

Akrich (2014, p. 180-181) aponta que os objetos técnicos são vistos como “instrumentos politicamente poderosos” porque impõem mediações que podem mudar as relações sociais. “A partir do momento em que o objeto técnico é quase estabilizado, ele torna-se, pelo seu desaparecimento, um instrumento de conhecimento”. A estabilização é obtida quando há uma adaptação mútua entre um objeto técnico e outros atores da rede em que é introduzido e ocorre “pela transformação do objeto técnico em caixa preta: ele desaparece ao mesmo tempo em que é mais indispensável que nunca”. Como menciona Akrich (2014, p. 181), essa relação estabilizadora ocorre quando outros campos são eliminados, “a economia se coloca à disposição dos objetos técnicos da mesma forma que a técnica se coloca à disposição da economia ou do social” (AKRICH, 2014).

É nesse sentido que os objetos técnicos podem ser considerados como os instrumentos politicamente poderosos: ao mesmo tempo em que eles produzem os modos de organização social, eles os naturalizam, os despolitizam, dá-lhes um conteúdo diferente (AKRICH, 2014, p. 181).

Akrich (1997) argumenta que os objetos técnicos fazem parte da construção de redes heterogêneas que ligam uma variedade de atores humanos e não humanos. Porque os objetos técnicos incorporam e mediam relações entre outros atores que formam “longas cadeias de pessoas, produtos, ferramentas, máquinas, dinheiro, etc.” (AKRICH, 1997, p. 205, tradução nossa). Neste contexto, elementos técnicos, sociais, econômicos etc. confluem nessas relações que tornam possível ser explicados da forma como as mesmas são estruturadas (AKRICH, 1997).

Em resposta à descrição dos papéis específicos que os objetos técnicos desempenham nas redes, Akrich (1997) propõe abandonar o determinismo tecnológico e o construtivismo social. Porque o determinismo tecnológico não dá atenção às relações mediadas e as reduz. O construtivismo social, por outro lado, nega que os objetos tomem o lugar dos atores. Concluindo que não se trata de

definir a tecnologia como instrumento de progresso ou método de submissão. E sim encontrar formas de examinar as relações formadas na sociedade (AKRICH, 1997).

Essas relações são construídas, mantidas e estabilizadas à medida que vários atores se movimentam constantemente entre o técnico e o social, e dentro e fora do objeto técnico. Há dois fatores a serem considerados ao lidar com objetos técnicos em rede. A primeira é questionar sua composição e mediação, no que se refere a atores humanos e não humanos. A segunda é observar a natureza desses atores e sua relevância quando transformam objetos e como são usados. A partir desse momento, os limites são estabelecidos levando em consideração as consequências de tais interações ao invés de algo simplesmente determinado. Essa fronteira é traçada entre o que um objeto técnico faz dentro da rede e sua função em relação a outros atores (AKRICH, 1997).

Por outro lado, a descrição desses objetos envolve dois problemas: metodologia e vocabulário. O vocabulário tende ser neutro e cuidadoso para evitar termos que impliquem uma distinção entre técnico e social. Já o problema metodológico é encontrar as situações internas e externas em que o objeto não está bem adaptado por meio de controvérsias, barganhas e potencial de dano em vez de apenas explicar os mecanismos elementares do funcionamento dos objetos técnicos (AKRICH, 1997).

Segundo Akrich (1991), seja a situação controversa ou não, descrever o que é o objeto técnico, como ele funciona e o ambiente, são abordagens que permitem a formulação de cenários ou roteiros para estabelecer uma rede sociotécnica. Pois, ao descobrir as associações produzidas pelas interações dos objetos técnicos entre os diversos atores envolvidos na rede representam um ou mais cenários que descrevem possíveis relações entre objetos técnicos e seu ambiente.

Em outras palavras, em situações de controvérsia, a primeira coisa que se busca é descobrir o que é ou faz o objeto técnico, o que leva, em segunda instância, a uma descrição do ambiente deste objeto, enquanto que em circunstâncias normais, trata-se de determinar (no duplo sentido de descrição e vontade) o ambiente a fim de determinar o objeto (AKRICH, 1991, p. 346, tradução nossa).

Akrich (2006) define a relação entre sistemas técnicos e contextos que superam múltiplas fronteiras através dos chamados modelos que vão desde a organização até as representações do mundo natural e físico. Uma forma de enfrentar esse problema é combinar a técnica e a construção social. Isso deixa claro que o desenvolvimento de objetos técnicos também consiste em fatores sociais e não puramente técnicos racionais. Nesse modelo proposto pela autora, a tecnologia faz parte da construção social, portanto, as escolhas técnicas feitas durante o desenvolvimento do objeto, suas decisões sociais, formação, relacionamentos, crenças sociais, etc., caracteriza aspectos sociotécnicos e desempenham um papel importante no rastreamento da origem das formas que os objetos assumem. Essa construção ocorre porque os atores “navegam constantemente em águas turbulentas entre o social, o técnico, o econômico, etc” (AKRICH, 2006, p. 32, tradução nossa) e formam relações de acordo com as circunstâncias do momento de mediação ao misturarem com elementos de forma heterogênea.

O trabalho dos atores pode ser descrito como um processo de problematização, o que os leva a formular um problema e, ao fazê-lo, definir os atores envolvidos na resolução do problema e, finalmente, colocar seu projeto em uma posição de passagem obrigatória para todos esses atores (AKRICH, 2006, p. 37-38, tradução nossa).

Akrich (2006) argumenta que, em vez de estabilizar os objetos técnicos como bem sucedidos por meio de parâmetros econômicos, também se deve avaliar a estabilização de parâmetros técnicos, sociais e organizacionais. Neste sentido, é por meio da ação que se torna consensual entre o objeto técnico e os demais atores no ambiente que estabelece este parâmetro.

Além disso, a autora observa que, por meio dessa relação entre técnica e social, emerge uma descrição do microuniverso de cada ator, possibilitando relações heterogêneas com outros atores que atuam como interseções na rede. Essa ação é descrita pela circulação dos objetos nos pontos de passagem obrigatórios onde ocorrem a redução dessas mediações, e conseqüentemente a redução da rede em apenas dois pontos possíveis de passagem desses objetos na qual são "globalizados em uma equivalência geral da rede" (AKRICH, 2006, p.52, tradução

nossa). Essa ação não é redundância de rede pois, esse movimento participa do método de equivalência e tratamento simétrico no momento em que eles estão sendo redefinidos. Dessa forma, os atores buscam sistemas sociotécnicos para se estabilizar (AKRICH, 2006).

O conceito de redes sociotécnicas aprofundado por Akrich foi introduzido por Latour (2012) e refere-se à conexão mútua entre pessoas e objetos prontos para integrar novas ações. Como resultado, as redes sociotécnicas cobrem um vasto espaço de possibilidades que promovem a produção e a circulação do conhecimento (LATOURE, 2012).

Da mesma forma, grupos de atores híbridos, mediam a produção e disseminação do conhecimento, formação de concepções, arranjos sociais e associações. Assim, os fenômenos sociais nascem dessas associações. Pois segundo Latour (2000, 2001, 2012) a sociedade é um produto de várias conexões entre os atores e refere-se a essa propriedade da técnica como a recombinação da sociedade que forma a definição de redes sociotécnicas.

As técnicas são aquilo que acontece a ferramentas e atuantes não-humanos quando processados por uma organização que os extrai, recombina e socializa. Até as técnicas mais simples são sociotécnicas; até nesse nível primitivo de significado as formas de organização revelam-se inseparáveis dos gestos técnicos. (LATOURE, 2001, p.140)

Assim, quando nos dedicamos à investigação das interações que culminam na criação de conexões entre os diversos atores em uma rede, temos a oportunidade de discernir e compreender profundamente fenômenos sociais complexos. Esse entendimento é alcançado por meio da análise dos comportamentos associativos dos atores, bem como do acompanhamento detalhado das ações que eles empreendem e das associações que emergem na construção da rede sociotécnica. Esse enfoque se torna particularmente crucial quando voltamos nossa atenção para o estudo dos objetos técnicos que permeiam essas interações (AKRICH, 2006).

3 INTELIGÊNCIA ARTIFICIAL

A definição do conceito de inteligência artificial é um tema que continua a suscitar debates acalorados entre os principais especialistas da área. Essa controvérsia surge devido à intrincada e ampla natureza do campo de estudo, que torna a delimitação precisa de sua definição uma tarefa subjetiva e desafiadora. Russel e Norvig (2013) contribuíram significativamente para a discussão sobre a subjetividade inerente à definição de IA ao analisarem as várias interpretações existentes e sua relação com os processos cognitivos e comportamentais.

Conforme destacado por Russel e Norvig (2013), a IA é uma disciplina vasta e interdisciplinar que se concentra na concepção e desenvolvimento de sistemas computacionais capazes de desempenhar tarefas que normalmente exigem habilidades humanas. Isso abrange uma ampla gama de capacidades, incluindo percepção, raciocínio, aprendizado e tomada de decisões. No entanto, é importante ressaltar que diferentes autores têm adotado perspectivas diversas ao definir a IA, enfocando métricas de sucesso que podem se basear no desempenho humano ou na busca da racionalidade ideal.

Essa multiplicidade de abordagens na definição da IA pode ser observada na Tabela 1, que apresenta várias interpretações do conceito de IA propostas por eminentes estudiosos da área, como Russel e Norvig (2013). A diversidade de perspectivas reflete a complexidade inerente à IA e sua constante evolução como um campo de estudo e prática. Essas diferentes visões permitem que a IA seja abordada de várias maneiras, adaptando-se às necessidades e contextos específicos, o que, por sua vez, impulsiona a inovação contínua nesse domínio fascinante. Portanto, a definição de IA permanece dinâmica e em constante evolução, refletindo a natureza em constante transformação dessa disciplina emocionante.

QUADRO 1 - Algumas definições de IA

Pensando como um humano	Pensando racionalmente
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) máquinas com mentes, no sentido total e literal.” (Haugeland, 1985 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p> <p>“[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p>
Agindo como seres humanos	Agindo racionalmente
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole <i>et al.</i>, 1998 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p> <p>“AI... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998 <i>apud</i> RUSSEL; NORVIG, 2013, p.25)</p>

FONTE: Adaptado de Russel e Norvig (2013, p. 25)

Os autores Russel e Norvig (2013) discutem a controvérsia em torno da definição da IA, devido à sua natureza ampla e complexa como objeto de estudo, o que leva a uma definição subjetiva. Eles apresentam quatro estratégias históricas para o estudo da IA: abordagem centrada nos seres humanos, abordagem racionalista, abordagem baseada em sistemas físicos e abordagem baseada em modelos matemáticos, cada uma conduzida por grupos distintos e com métodos variados que ajudaram e desacreditaram uns aos outros ao longo do tempo.

Para Newquist (2020), previamente, a AI era definida convencionalmente como um conjunto de tecnologias que emulavam o comportamento humano. Cada

subcampo da IA, como sistemas especializados, processamento de linguagem natural, reconhecimento de voz, visão computacional e redes neurais, era considerado uma forma de IA, uma vez que imitava algum aspecto do comportamento humano. Ademais, o sucesso de qualquer uma dessas formas de IA no mundo dos negócios era tomado como prova de que a tecnologia era de fato inteligente. Entretanto, esse raciocínio não contribuiu para uma definição mais precisa da IA.

Contudo, conforme afirma Luger (2013), a IA é um campo da ciência da computação que se dedica à automatização de comportamento inteligente por meio de conhecimentos, algoritmos³ e técnicas de programação. No entanto, esta definição sofre do fato de que a própria inteligência é mal definida e compreendida. Embora muitas pessoas possam identificar comportamentos inteligentes, o conceito de inteligência em si não é claramente definido e compreendido. Portanto, criar uma definição precisa o suficiente para avaliar um programa de computador que se diz inteligente é um desafio (LUGER, 2013).

Após superar o desafio de criar uma "inteligência", Luger (2013) reitera que os pesquisadores de IA frequentemente atuam como engenheiros, projetando artefatos inteligentes específicos. Esses artefatos se manifestam como ferramentas de diagnóstico, prognóstico ou visualização, que oferecem assistência aos usuários humanos para a realização de tarefas complexas.

³ Os algoritmos desempenham um papel fundamental na matemática e ciência da computação, eles são conjuntos de instruções ou procedimentos bem definidos, geralmente na forma de uma sequência de passos, utilizados para resolver um problema ou executar uma tarefa específica. Em termos mais simples, um algoritmo pode ser comparado a uma receita de culinária. Assim como uma receita define uma sequência de passos para preparar um prato específico, um algoritmo define uma sequência de passos para resolver um problema. Esses passos podem envolver operações matemáticas, estruturas de controle, manipulação de dados e outras ações. Eles podem ser expressos em linguagens de programação específicas ou em pseudocódigo, que é uma forma mais próxima da linguagem humana (BLEAKLEY, 2020). Na história da programação, Ada Lovelace (1815-1852) é aclamada como pioneira por criar o primeiro algoritmo interpretado por uma máquina. Em colaboração com Charles Babbage, ela desenvolveu algoritmos para a máquina analítica de Babbage, precursora dos computadores modernos, permitindo a computação de funções matemáticas (FUEJI; FRANCIS, 2003). Na IA, os algoritmos são a base técnica que permite que os sistemas aprendam, tomem decisões e executem tarefas de maneira inteligente (RUSSEL; NORVIG, 2013).

O problema de definir o campo inteiro da inteligência artificial é semelhante ao definir a própria inteligência: ela é uma única faculdade ou é apenas um nome para uma coleção de capacidades distintas e não relacionadas? Até que ponto a inteligência é aprendida e não existe desde o nascimento? O que acontece exatamente quando ocorre o aprendizado? O que é criatividade? O que é intuição? A inteligência pode ser deduzida do comportamento observável ou ela requer evidências de um mecanismo interno em particular? Como o conhecimento é representado no tecido nervoso de um ser humano e que lições isso nos traz para o projeto de máquinas inteligentes? O que é autopercepção? Que papel ela desempenha na inteligência? Além disso, o conhecimento sobre a inteligência humana é necessário para construir um programa inteligente, ou uma técnica estritamente de "engenharia" é suficiente para tratar do problema? É possível conseguir inteligência em um computador, ou uma entidade inteligente requer a riqueza de sensações e experiências que só poderiam ser encontradas em uma existência biológica? (LUGER, 2013, p. 1-2).

Em virtude desses fatores, uma definição exata de IA pode ser considerada ambígua. Na verdade, segundo Luger (2013), essa definição tem gerado mais perguntas e resultou em uma noção paradoxal de um campo de estudo que busca, dentre outras coisas, definir a si mesmo.

Conforme alegado por Luger (2013), a busca por responder às questões acima mencionadas tiveram um impacto notável na formulação de problemas e metodologias de solução que constituem o cerne da IA moderna. De fato, uma das características distintivas da IA reside em sua singular e poderosa habilidade para explorar precisamente tais questões. Pois a IA fornece um meio e um campo de testes para teorias de inteligência, as quais podem ser expressas em linguagem de programação de computadores e, assim, ser testadas e validadas por meio da execução desses programas em um computador.

Já que tradicionalmente, segundo Luger (2013), o campo de estudo da IA concentrou-se mais na expansão das capacidades da ciência da computação do que na definição de seus limites. O autor argumenta que é crucial assegurar que essa expansão seja fundamentada em princípios teóricos sólidos da área da computação

e estabelece que a IA é um empreendimento humano e sua definição é composta pela coleção de problemas e metodologias estudadas pelos pesquisadores da área. Ademais, o próprio conceito de IA tem evoluído ao longo do tempo à medida que novas tecnologias e abordagens são desenvolvidas.

Em virtude de seu escopo e da sua ambição, a inteligência artificial não tem uma definição simples. Até o momento, simplesmente a definimos como a coleção de problemas e metodologias estudada pelos pesquisadores de inteligência artificial. Essa definição pode parecer tola e sem sentido, mas ela reforça um argumento importante: a inteligência artificial, como toda ciência, é um empreendimento humano e talvez seja mais bem entendida nesse contexto. Há várias razões para que qualquer ciência, incluindo a IA, se preocupe com um certo conjunto de problemas e desenvolva um conjunto de técnicas específicas para enfrentar esses problemas. (LUGER, 2013, p. 2).

As classificações de IA também são questões controversas e ainda em debate na comunidade. Existem várias formas de classificar a IA, cada uma delas enfatizando diferentes aspectos e características da tecnologia. A adequada classificação da IA dependerá do objetivo específico e do contexto em que a mesma será aplicada. Inicialmente, uma forma comum de classificar a IA era por seu nível de "inteligência" (RUSSEL; NORVIG, 2013).

O artigo "Mentes, Cérebros e Programas" (*Minds, Brains, and Programs*) de John Searle é um ensaio filosófico que discute a possibilidade de máquinas pensarem e terem consciência. Publicado em 1980, o artigo foi uma resposta às teorias da IA que afirmavam que era possível criar máquinas capazes de pensar e ter consciência. O autor questiona a possibilidade de que uma máquina possa realmente "entender" ou ter consciência de algo, mesmo que ela possa simular comportamentos humanos inteligentes (SEARLE, 1980) (MITCHELL, 2019).

Para Searle (1980), a compreensão de algo requer uma compreensão do significado das palavras usadas para descrever o objeto em questão. E, segundo ele, uma máquina não pode realmente entender o significado das palavras, porque

ela só é capaz de manipular símbolos de forma mecânica, sem compreender o significado real que eles representam.

O autor, portanto, apresenta o experimento mental do Quarto Chinês⁴ para ilustrar seu argumento de que uma máquina que opera por meio de símbolos e regras pré-programadas não pode ter realmente entendimento ou consciência, independentemente de sua capacidade de simular comportamentos humanos inteligentes. Em outras palavras, a IA pode ser capaz de realizar tarefas com grande eficiência, mas isso não significa que ela tenha compreensão ou consciência real (SEARLE, 1980).

O artigo de Searle (1980) é relevante por ter classificado a IA em duas categorias principais: a IA forte (*strong AI*) e a IA fraca (*weak AI*). A IA forte, também conhecida como IA geral, é uma forma hipotética de IA capaz de emular ou superar a inteligência humana em todas as áreas, incluindo habilidades como aprendizado, raciocínio, resolução de problemas, compreensão de linguagem natural e criatividade. Segundo Searle (1980), uma IA forte seria capaz de possuir uma mente e, portanto, capaz de ter consciência e experiência subjetiva (MITCHELL, 2019).

Searle (1980) define a IA fraca como uma forma de IA projetada para desempenhar tarefas específicas de maneira mais eficiente do que um ser humano. Essas tarefas geralmente envolvem a análise de grandes quantidades de dados e a tomada de decisões com base nessas análises. Exemplos de IA fraca incluem sistemas de reconhecimento de fala, sistemas de recomendação de produtos, sistemas de diagnóstico médico e assistentes virtuais (LEE, 2018) (MITCHELL, 2019). Searle (1980) argumenta que a IA fraca não é capaz de possuir uma mente e, portanto, não é capaz de ter consciência ou experiência subjetiva. Para ele, a IA fraca é apenas uma simulação da inteligência humana e não é capaz de pensar ou compreender como um ser humano.

⁴No experimento do "Quarto Chinês", uma pessoa que não fala chinês é colocada em um quarto com símbolos chineses, um manual em inglês e um sistema de entrada e saída de cartões. Ela manipula os símbolos com base nas instruções do manual, produzindo respostas em chinês. No entanto, isso não implica que a pessoa realmente compreenda o chinês; é apenas o processamento de informações de acordo com regras predefinidas (SEARLE, 1980). Este experimento, frequentemente utilizado para discutir a inteligência artificial, destaca a diferença entre a capacidade de processamento de informações de um sistema, como um computador, e a verdadeira compreensão ou consciência (MITCHELL, 2019).

A distinção entre IA forte e fraca é uma questão fundamental dentro da área de IA. A IA fraca é atualmente a forma mais comum de IA e útil em diversas áreas. Enquanto a IA forte ainda é vista como um objetivo a ser alcançado no futuro. Apesar de a IA forte ainda ser vista como um objetivo a ser alcançado no futuro, muitos pesquisadores e cientistas acreditam que ela pode representar um grande avanço na história da humanidade (LEE, 2018) (MITCHELL, 2019).

O artigo de Searle (1980) gerou muitos debates e controvérsias na área da IA. Alguns argumentam que suas conclusões são simplistas ou equivocadas, enquanto outros concordam com a ideia de que a compreensão real requer algo mais do que apenas processamento de símbolos. No entanto, é inegável que o trabalho de Searle teve um impacto significativo no desenvolvimento da teoria da IA e na forma como as pessoas pensam sobre a natureza da consciência e do entendimento humano (MITCHELL, 2019).

Entretanto, desde 2015, os pesquisadores têm se tornado mais precisos nas definições modernas de IA, embora ainda permitam certa flexibilidade. Essas definições são classificadas como "IA estreita" (*Narrow AI*) e "IA geral" (*general / board AI*) (NEWQUIST, 2020). Há pesquisadores e desenvolvedores que buscam criar sistemas que possam realizar tarefas de forma inteligente, como um ser humano, enquanto outros não consideram a funcionalidade humana como essencial, desde que o sistema cumpra sua finalidade corretamente. Além dessas abordagens, há outras que se encontram em algum ponto intermediário, utilizando o raciocínio humano como base para orientar a forma como os computadores podem realizar tarefas semelhantes (HAMMOND, 2015).

A IA pode ser programada para desempenhar tarefas específicas, conhecida como IA estreita, ou para ter capacidade de raciocínio geral, chamada de IA ampla ou geral. A aplicação comercial da IA estreita é mais evidente, pois se concentra em resolver problemas específicos, como um sistema de reconhecimento de fala que pode ter um bom desempenho em um idioma específico, mas não em outro. No entanto, a IA estreita apresenta a desvantagem de ter limitações na realização de tarefas fora de seu escopo. É pouco provável que se deseje utilizar um sistema altamente eficiente em encontrar postos de gasolina mais próximos para realizar diagnósticos médicos. Apesar de haver pessoas que não classificam sistemas de IA

estreita como verdadeira IA, esses sistemas são amplamente empregados em aplicações comerciais. Um exemplo é quando a Amazon faz recomendações de produtos baseadas no histórico de compras do usuário, o que demonstra um sistema de IA estreita em funcionamento (HAMMOND, 2015).

A IA estreita é tudo o que a pesquisa de IA tem feito até hoje. Ele define um programa que pode fazer uma coisa (ou várias coisas relacionadas) muito bem. É isso, no entanto. Como os sistemas especialistas de antigamente, a IA estreita descreve um sistema que pode, por exemplo, detectar padrões estranhos nos gastos com cartão de crédito – mas não consegue entender a fala e não consegue encontrar a rota mais rápida de sua casa até o Burger King mais próximo. Cada uma dessas habilidades requer seu próprio aplicativo de IA específico (NEWQUIST, 2020, p. 665, tradução nossa).

Em contrapartida, a IA geral é desenvolvida para ter capacidade de raciocínio geral e pode ser utilizada em diversas tarefas. Entretanto, o desenvolvimento de sistemas de IA ampla ainda se encontra em estágio experimental, havendo poucos exemplos comerciais conhecidos, como o aprendizado profundo do Google (desenvolvido para categorizar automaticamente a partir de exemplos) e o Watson da IBM (desenvolvido para inferir conclusões a partir de grandes quantidades de evidências textuais) (HAMMOND, 2015).

A IA geral é tudo o que a ficção científica postulou e previu para máquinas inteligentes nas últimas sete décadas. Uma máquina que pode pensar, reagir, aprender, adaptar e se comportar como um ser humano. Na verdade, a IA geral é o objetivo de todas as pesquisas sobre IA desde o início do Teste de Turing: uma máquina cuja atividade intelectual é indistinguível da de um humano (NEWQUIST, 2020, p. 665, tradução nossa).

Para Newquist (2020), durante a segunda metade e o final do século 20, a maioria dos avanços no campo da IA foi direcionada para o desenvolvimento de sistemas de IA estreita, embora tenha sido realizado com a suposição de que estaria contribuindo para a IA geral. Os pesquisadores acreditavam estar contribuindo para o avanço da IA geral, mas, na realidade, estavam limitando rigorosamente cada aplicação da IA estreita, resultando no seu progressivo estreitamento.

3.1 INTELIGÊNCIA ARTIFICIAL MODERNA

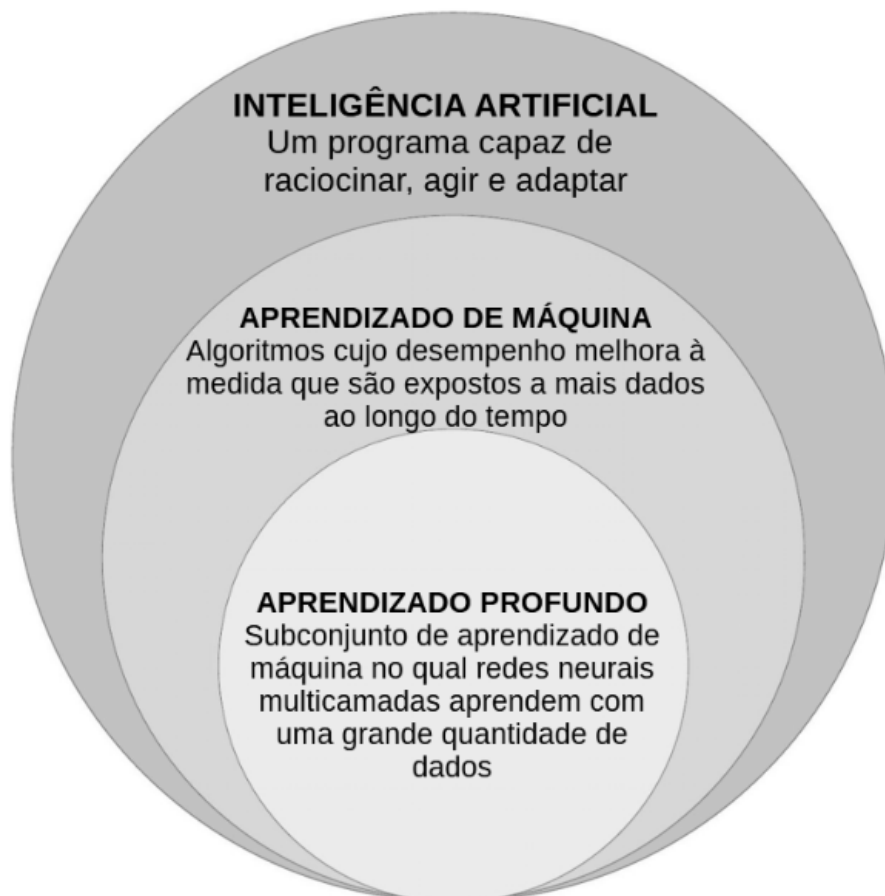
A IA moderna se refere ao estado atual da pesquisa e desenvolvimento da IA. Ela se apoia em avanços tecnológicos e metodológicos recentes que impulsionam o progresso nessa área. Na IA moderna, os sistemas são projetados para adquirir conhecimento a partir dos dados disponíveis e tomar decisões ou realizar tarefas com base nesse aprendizado. Em vez de serem programados de forma explícita para executar uma tarefa específica, os sistemas de IA modernos são treinados por meio de algoritmos de aprendizado de máquina e grandes conjuntos de dados (TAULLI, 2019). Alguns dos principais avanços na IA moderna incluem:

- **Aprendizado de máquina (*Machine Learning*):** Algoritmos de aprendizado de máquina, como redes neurais artificiais, são usados para treinar sistemas de IA para reconhecer padrões complexos nos dados e realizar tarefas específicas (TAULLI, 2019).
- **Aprendizado profundo (*Deep Learning*):** É uma subárea do aprendizado de máquina que se concentra em redes neurais artificiais com várias camadas, capazes de aprender representações de dados em vários níveis de abstração. O aprendizado profundo tem sido especialmente eficaz no processamento de imagens, reconhecimento de fala e tradução automática (TAULLI, 2019).
- **Processamento de linguagem natural (*Natural Language Processing - NLP*):** Envolve o desenvolvimento de algoritmos e modelos que permitem que os computadores entendam, interpretem e gerem a linguagem humana de maneira natural. O NLP tem sido aplicado em chatbots, assistentes virtuais e sistemas de tradução automática (RUSSEL; NORVIG, 2013) (TAULLI, 2019).
- **Visão computacional:** Consiste em capacitar os computadores a interpretar e entender informações visuais, semelhante à maneira como os seres humanos processam as imagens. Isso tem aplicações em reconhecimento facial, classificação de imagens, veículos autônomos e muito mais (RUSSEL; NORVIG, 2013) (TAULLI, 2019).

→ Robótica e automação: A IA moderna também tem sido aplicada em robótica e automação, permitindo que os robôs e outros dispositivos autônomos realizem tarefas complexas e interajam com o ambiente de maneira mais inteligente (RUSSEL; NORVIG, 2013) (TAULLI, 2019).

A fim de compreender a estrutura da IA moderna, a Figura 1 ilustra como os principais componentes da IA estão interconectados. No topo da hierarquia encontra-se a própria IA, a qual engloba uma vasta gama de teorias e tecnologias. Essa área pode ser subdividida em duas categorias principais: aprendizado de máquina e aprendizado profundo (TAULLI, 2019).

FIGURA 1 - Uma perspectiva sobre os elementos fundamentais da IA Moderna



Fonte: Adaptado de Alzubaidi *et al.* (2021, p. 7, tradução nossa)

É importante destacar que a IA moderna ainda está em constante evolução, e os pesquisadores e engenheiros continuam a explorar novas técnicas e abordagens para melhorar sua eficácia e aplicabilidade em diversos setores.

3.1.1 Aprendizado de máquina

O aprendizado de máquina (*machine learning*), é uma abordagem que permite treinar um computador para fazer previsões precisas por meio da análise de dados. Ao contrário do desenvolvimento tradicional de software, que se baseia em sequências lógicas de comandos, o aprendizado de máquina capacita o computador a identificar padrões e relações nos dados fornecidos. Essa capacidade de identificação de padrões permite que o computador tome decisões cada vez mais precisas com base em princípios estatísticos (TAULLI, 2019).

Essas estatísticas incluem o desvio padrão, que mede a dispersão dos dados em relação à média, a distribuição normal, que descreve a distribuição probabilística dos dados, o teorema de Bayes, que permite atualizar probabilidades com base em novas evidências, a correlação, que mensura o grau de relação entre duas variáveis, e a extração de características, que envolve identificar as informações mais relevantes e distintivas nos dados para melhorar o desempenho do modelo de aprendizado de máquina (TAULLI, 2019).

Os conceitos estatísticos desempenham um papel crucial no desenvolvimento e compreensão dos algoritmos de aprendizado de máquina, sendo essenciais para alcançar resultados precisos e confiáveis, pois os modelos podem generalizar esse conhecimento para realizar previsões ou tomar decisões quando expostos a novos dados (TAULLI, 2019).

Segundo Hurwitz e Kirsch (2018), o processo de aprendizado de máquina envolve três etapas principais:

1. **Representação:** Durante essa etapa, um modelo de aprendizado de máquina é desenvolvido com base em um conjunto de dados de treinamento. Esse conjunto de dados contém exemplos de entrada (conhecidas como "amostras") e as saídas correspondentes (conhecidas como "rótulos" ou "alvos"). O modelo analisa esses dados de treinamento e busca padrões ou relações entre as amostras e os rótulos. Existem diferentes tipos de algoritmos de aprendizado de máquina, como regressão linear, árvores de decisão, redes neurais, entre outros, cada um com suas próprias características e adequado para diferentes tipos de problemas.
2. **Teste/Validação:** Após o treinamento, o modelo precisa ser avaliado para verificar sua eficácia. Para isso, utiliza-se um conjunto de dados de teste ou validação que não foi visto durante o treinamento. O modelo faz previsões com base nessas amostras de teste e os resultados são comparados com os rótulos reais. Métricas de desempenho, como precisão, recall ou erro médio, são calculadas para avaliar a qualidade das previsões do modelo.
3. **Otimização:** Depois de treinado e validado, o modelo é capaz de fazer previsões ou tomar decisões com base em novos dados de entrada. O modelo aplica os conhecimentos adquiridos durante o treinamento para fornecer uma saída ou uma ação com base nas entradas recebidas. Essa capacidade de generalização é uma das principais características do aprendizado de máquina.

Para que o processo de aprendizado de máquina seja bem-sucedido, é fundamental ter um conjunto de dados de treinamento representativo e de qualidade, bem como escolher o algoritmo adequado para o problema em questão. Além disso, é importante realizar ajustes no modelo, como otimização de hiperparâmetros, para obter melhores resultados (HURWITZ; KIRSCH, 2018).

Existem vários tipos de técnicas de aprendizado no aprendizado de máquina, cada um com abordagens e técnicas específicas. Segundo Sarker (2021), os principais tipos de aprendizado são divididos em quatro categorias:

- **Aprendizado Supervisionado:** Nesse tipo de aprendizado, o modelo é treinado usando um conjunto de dados de treinamento que contém exemplos rotulados, ou seja, pares de entrada e saída esperada. O objetivo é aprender a mapear as entradas para as saídas corretas. Por exemplo, um modelo de classificação pode ser treinado para reconhecer imagens de gatos e cães com base em um conjunto de imagens rotuladas. O modelo aprende a distinguir as características que diferenciam os gatos dos cães e, em seguida, pode classificar imagens não vistas anteriormente .
- **Aprendizado Não Supervisionado:** Nesse tipo de aprendizado, o modelo é treinado em um conjunto de dados não rotulados, onde não há informações sobre as saídas esperadas. O objetivo é descobrir padrões, estruturas ou relações intrínsecas nos dados. Algoritmos de clusterização são um exemplo comum de aprendizado não supervisionado, onde o objetivo é agrupar os dados em clusters com base em suas similaridades.
- **Aprendizado por Reforço:** Nesse tipo de aprendizado, um agente aprende a tomar decisões sequenciais em um ambiente para maximizar uma recompensa cumulativa. O agente interage com o ambiente, recebe feedback em forma de recompensas ou penalidades e, ao longo do tempo, aprende a tomar as ações que levam às maiores recompensas. Esse tipo de aprendizado é frequentemente usado em jogos, robótica e otimização de processos.
- **Aprendizado Semi-Supervisionado:** Nesse tipo de aprendizado, o modelo é treinado com um conjunto de dados que contém tanto exemplos rotulados quanto não rotulados. O aprendizado semi-supervisionado visa aproveitar os dados rotulados disponíveis, juntamente com a estrutura geral dos dados não rotulados, para melhorar o desempenho do modelo.

Esses são alguns dos principais tipos de técnicas de aprendizado no aprendizado de máquina, cada um com suas técnicas e abordagens específicas.

Dependendo do problema e dos dados disponíveis, diferentes tipos de aprendizado podem ser aplicados para obter os melhores resultados (SARKER, 2021).

3.1.2 Aprendizado profundo

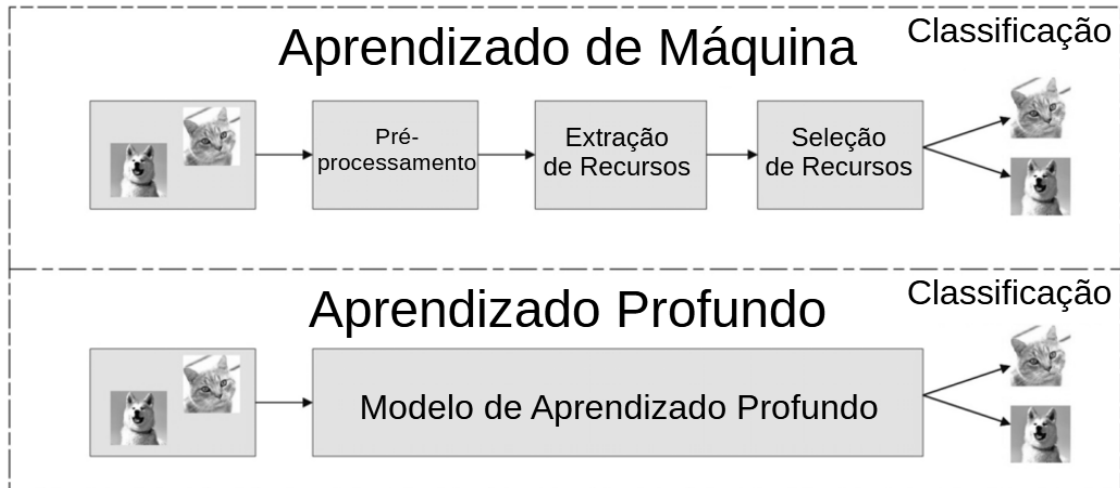
Recentemente, o paradigma de aprendizado profundo tem se destacado como o método predominante na comunidade de aprendizado de máquina. Gradualmente, tem se tornado a abordagem computacional mais amplamente empregada no campo do aprendizado de máquina, demonstrando resultados notáveis em uma variedade de tarefas cognitivas complexas, alcançando e, em alguns casos, até mesmo superando o desempenho humano (ALZUBAIDI *et al.*, 2021).

Uma das principais vantagens do aprendizado profundo reside em sua capacidade de lidar com grandes volumes de dados. Nos últimos anos, esse campo tem experimentado um crescimento acelerado e tem sido extensivamente utilizado para atender com sucesso a uma ampla gama de aplicações tradicionais. É importante ressaltar que o aprendizado profundo tem superado as técnicas conhecidas de aprendizado de máquina em diversos domínios, como segurança cibernética, processamento de linguagem natural, bioinformática, robótica, controle e processamento de informações médicas, entre outros (ALZUBAIDI *et al.*, 2021).

O aprendizado de máquina refere-se ao campo de estudo e prática no qual os sistemas de computador são projetados para aprender e melhorar a partir de dados sem serem explicitamente programados. Envolve o desenvolvimento de algoritmos e modelos que podem analisar e interpretar dados, identificar padrões e fazer previsões ou decisões com base nas informações coletadas (LOWE, 2018).

O aprendizado profundo, por outro lado, é um subconjunto do aprendizado de máquina que se concentra no desenvolvimento e uso de redes neurais artificiais, particularmente redes neurais profundas. Essas redes são projetadas para simular o comportamento dos neurônios no cérebro humano, permitindo que eles reconheçam padrões e extraiam representações significativas de dados complexos (LOWE, 2018).

FIGURA 2 - A diferença entre aprendizado profundo e aprendizado de máquina



Fonte: Adaptado de Alzubaidi *et al.* (2021, p. 7, tradução nossa)

Uma das principais características do aprendizado profundo é a capacidade de criar modelos hierárquicos com várias camadas de nós ou neurônios interconectados. Cada camada na rede processa e extrai recursos cada vez mais abstratos e complexos dos dados de entrada. Essa representação hierárquica permite que os modelos de aprendizado profundo capturem relações e dependências intrincadas nos dados, levando a previsões ou classificações mais precisas e diferenciadas (LOWE, 2018).

Em comparação com as técnicas tradicionais de aprendizado de máquina, os algoritmos de aprendizado profundo demonstraram desempenho superior em uma ampla gama de aplicações, incluindo visão computacional, processamento de linguagem natural, reconhecimento de fala e sistemas de recomendação, entre outros. Ao alavancar grandes quantidades de dados e recursos computacionais poderosos, os modelos de aprendizado profundo podem aprender e generalizar a partir de padrões complexos e fazer previsões ou decisões altamente precisas (LOWE, 2018).

O impacto do aprendizado profundo pode ser visto em vários aspectos de nossas vidas diárias. Os mecanismos de recomendação, como os usados pela

Netflix e Amazon, dependem de sistemas de aprendizado profundo para analisar as preferências e o comportamento do usuário, bem como grandes quantidades de dados sobre produtos, filmes ou programas. Esses sistemas podem fornecer recomendações personalizadas que correspondam aos interesses dos usuários, levando a uma experiência de usuário mais satisfatória e maior envolvimento (LOWE, 2018).

Além disso, o aprendizado profundo revolucionou as tarefas de visão computacional. Ele permitiu avanços significativos no reconhecimento de imagens e objetos, permitindo que os sistemas identifiquem e classifiquem com precisão objetos em imagens ou vídeos. Isso encontrou aplicações em áreas como carros autônomos, sistemas de vigilância, imagens médicas e reconhecimento facial (LOWE, 2018).

O processamento de linguagem natural também se beneficiou de técnicas de aprendizado profundo. Modelos de aprendizado profundo, como redes neurais recorrentes e modelos transformadores, alcançaram resultados impressionantes em tarefas como tradução de idiomas, análise de sentimentos, geração de texto e chatbots. Esses avanços melhoraram a qualidade dos serviços de tradução automática, sistemas automatizados de suporte ao cliente e aplicativos de compreensão de idiomas (LOWE, 2018).

No geral, o aprendizado profundo representa um salto significativo nas capacidades dos sistemas de aprendizado de máquina. Ao incorporar redes neurais com várias camadas, os modelos de aprendizado profundo podem aprender representações de dados mais complexas e obter desempenho de ponta em vários domínios. Seu impacto em sistemas de recomendação, visão computacional, processamento de linguagem natural e outras áreas continua a crescer, moldando a forma como interagimos com a tecnologia e melhorando a qualidade de nossas vidas (LOWE, 2018).

Em síntese, um cientista de dados utilizando IA baseada em regras e aprendizado de máquina desempenha o papel primordial de realizar a seleção de regras e características a serem empregadas nos modelos, estabelecendo, assim, o funcionamento desses modelos. Por outro lado, no contexto do aprendizado

profundo, os cientistas de dados encaminham dados brutos para um algoritmo, o qual analisa esses dados e, com base no conhecimento acumulado e nas inferências derivadas dos novos dados apresentados, efetua previsões (LOWE, 2018).

As redes neurais artificiais são sistemas de computação que se baseiam no funcionamento do cérebro humano. Esses sistemas são compostos por elementos de processamento interconectados, conhecidos como neurônios artificiais, que colaboram para executar atividades de aprendizado e identificação de padrões (GOODFELLOW; BENGIO; COURVILLE, 2016).

Uma rede neural profunda, como uma estrutura de aprendizado profundo, pode ser composta por várias camadas (*layers*) de unidades de processamento. Cada camada tem um papel específico no processamento dos dados de entrada e na produção dos resultados finais (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

A primeira camada, também conhecida como camada de entrada, recebe os dados de entrada, que podem ser diferentes tipos de informações, como imagens, texto ou sequências numéricas. A camada de entrada geralmente não realiza nenhum processamento nos dados, mas os envia para as camadas intermediárias (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

As camadas intermediárias, também chamadas de camadas ocultas, são responsáveis pelo processamento dos dados. Cada camada intermediária contém várias unidades de processamento, também chamadas de neurônios artificiais. Cada neurônio em uma camada oculta recebe informações da camada anterior, realiza um cálculo ponderado dessas informações usando pesos associados às conexões e aplica uma função de ativação. A função de ativação introduz não-linearidade na rede neural, permitindo que ela aprenda padrões e relações complexas nos dados. O resultado calculado em cada neurônio é então transmitido para os neurônios da próxima camada (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

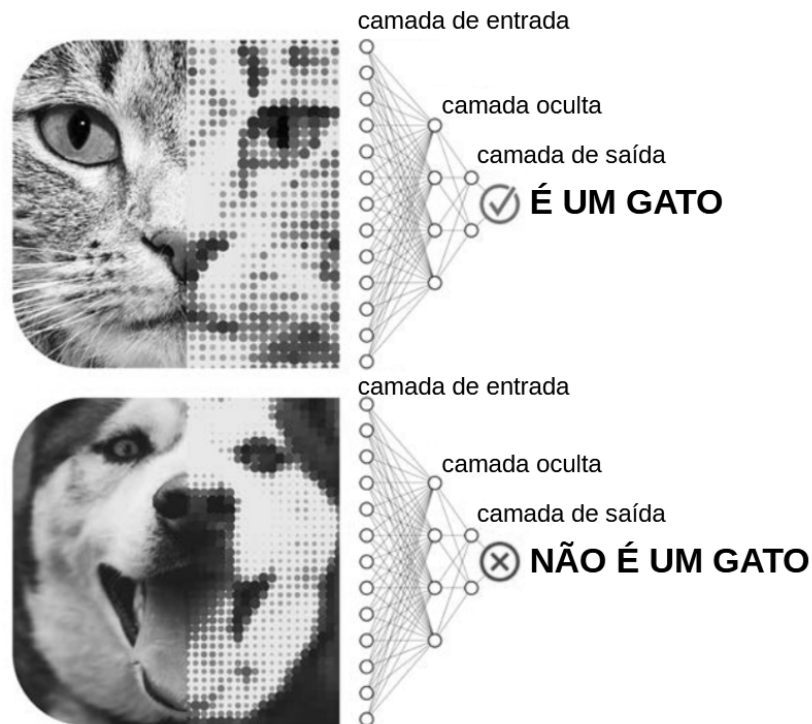
A camada de saída é a última camada da rede neural e é responsável por produzir os resultados finais. Cada neurônio na camada de saída representa uma

classe ou um valor numérico específico, dependendo da tarefa realizada pela rede neural. Por exemplo, em uma rede neural usada para classificar imagens, cada neurônio de saída pode representar uma classe diferente, como "cachorro", "gato" ou "carro". Durante o treinamento, os pesos das conexões da camada de saída são ajustados para minimizar a diferença entre os resultados previstos pela rede neural e os valores corretos (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

Treinar uma rede neural envolve expor a rede a um conjunto de dados de treinamento, no qual cada exemplo contém uma entrada e o valor correto correspondente. A rede neural ajusta de forma iterativa os pesos das conexões em todas as camadas para minimizar uma função de perda, que avalia o quão bem os resultados previstos pela rede se aproximam dos valores corretos. Esse processo de ajuste de pesos é realizado por meio de algoritmos de otimização, como o gradiente descendente, que calculam a direção na qual os pesos devem ser atualizados para reduzir a perda (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

Durante o processo de treinamento, a rede neural adquire a habilidade de identificar características importantes dos dados de entrada e de fazer previsões ou classificações precisas com base nessas características. Quanto mais camadas a rede neural tiver, maior será sua capacidade de aprender representações dos dados em níveis hierárquicos e complexos. No entanto, o treinamento de redes neurais profundas pode ser um desafio, pois requer uma quantidade considerável de dados e poder computacional para ajustar eficientemente os milhões de parâmetros envolvidos nessas redes (GOODFELLOW; BENGIO; COURVILLE, 2016) (BOUCHER, 2020).

FIGURA 3 - Esquema de uma rede neural para reconhecer imagens de gatos



Fonte: Adaptado de Boucher (2020, p. 4, tradução nossa)

Existem vários tipos de redes neurais, cada uma com sua própria arquitetura e características específicas. Alguns dos principais tipos de redes neurais destacadas por Alzubaidi *et al.* (2021) são:

- **Redes Neurais Feedforward:** Também conhecidas como perceptrons de múltiplas camadas (Multilayer Perceptrons ou MLPs), são redes em que a informação flui em uma direção, da camada de entrada para a camada de saída. Elas são compostas por uma camada de entrada, uma ou mais camadas intermediárias (camadas ocultas) e uma camada de saída. As redes feedforward são amplamente utilizadas em problemas de classificação, regressão e reconhecimento de padrões (GOODFELLOW; BENGIO; COURVILLE, 2016) (ALZUBAIDI *et al.*, 2021).
- **Redes Neurais Recorrentes (RNNs):** São redes que possuem conexões retroalimentadas, o que permite que informações anteriores sejam mantidas e usadas para processar entradas sequenciais. Essa capacidade de memória

torna as RNNs adequadas para lidar com tarefas como previsão de séries temporais, tradução de idiomas e processamento de linguagem natural (GOODFELLOW; BENGIO; COURVILLE, 2016) (ALZUBAIDI *et al.*, 2021).

- Segundo Taulli (2019), existem algumas desvantagens associadas às RNNs. Uma delas é conhecida como gradiente de fuga, no qual a precisão dos modelos diminui à medida que eles se tornam mais complexos. Além disso, esses modelos podem exigir um tempo maior para o treinamento. No entanto, o Google desenvolveu um modelo Transformer, que processa entradas em paralelo e alcança resultados mais precisos. Atualmente, existem arquiteturas aprimoradas do Transformer, como GPT-3, GPT-4 e o *chatbot* ChatGPT da OpenAI, além de outros modelos em desenvolvimento por diferentes empresas (CHEN *et al.*, 2023).
- Redes Neurais Convolucionais (CNNs): São redes especializadas no processamento de dados com estrutura de grade, como imagens. As CNNs são compostas por camadas convolucionais, que extraem características relevantes das imagens, e camadas de pooling, que reduzem a dimensionalidade dos dados. Essas redes são amplamente utilizadas em tarefas de classificação e reconhecimento de imagens (GOODFELLOW; BENGIO; COURVILLE, 2016) (ALZUBAIDI *et al.*, 2021).
- Redes Neurais Generativas Adversariais (GANs): São redes compostas por dois modelos: um gerador e um discriminador. O gerador cria amostras sintéticas a partir de um ruído aleatório, enquanto o discriminador avalia se essas amostras são reais ou falsas. Através de um processo iterativo de treinamento, o gerador e o discriminador se aprimoram mutuamente, resultando em amostras cada vez mais realistas. As GANs são usadas para gerar imagens sintéticas, sintetizar dados e realizar outras tarefas relacionadas à geração de conteúdo (GOODFELLOW; BENGIO; COURVILLE, 2016) (ALZUBAIDI *et al.*, 2021).

Há uma diversidade de redes neurais disponíveis, e os mencionados anteriormente são apenas alguns exemplos. Além desses, existem outras variações e estruturas desenvolvidas especificamente para atender a diversas necessidades e solucionar problemas específicos (GOODFELLOW; BENGIO; COURVILLE, 2016).

4 A EVOLUÇÃO DA INTELIGÊNCIA ARTIFICIAL

A história IA remonta a mais de meio século atrás. A ideia de criar máquinas capazes de pensar e tomar decisões como seres humanos é um conceito que tem sido objeto de especulação e fantasia há séculos. No entanto, foi apenas durante a década de 1950 que a IA começou a se tornar uma realidade tangível. O campo de estudo acadêmico da IA teve início com John McCarthy, que se interessou por computadores em 1948, após participar de um seminário intitulado "Mecanismos Cerebrais no Comportamento", que discutiu sobre a capacidade das máquinas de pensar no futuro. Dentre os participantes do seminário estavam grandes pioneiros na área, como John von Neumann, Alan Turing e Claude Shannon (TAULLI, 2019).

Em 1956, McCarthy então, liderou a realização de uma Conferência de Verão em Dartmouth College, NH, EUA. Durante o evento, o termo "Inteligência Artificial" foi cunhado por ele, Marvin Minsky (Harvard), Nathaniel Rochester (IBM) e Claude Shannon (Bell), que submeteram uma proposta intitulada "*Uma proposta para o Projeto de Pesquisa de Verão de Dartmouth sobre Inteligência Artificial, 31 de agosto de 1955*" (*A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*) à fundação Rockefeller. A proposta visava a realização de um estudo de dois meses com uma equipe de dez homens sobre o tópico da IA, sendo a primeira menção oficial à expressão (MCCORDUCK, 2004) (RICH; KNIGHT; NAIR, 2010).

Propomos que um estudo de inteligência artificial de 2 meses e 10 homens seja realizado durante o verão de 1956 no Dartmouth College em Hanover, New Hampshire. O estudo deve prosseguir com base na conjectura de que todo aspecto do aprendizado ou qualquer outra característica da inteligência pode, em princípio, ser descrito com tanta precisão que uma máquina pode ser feita para simulá-lo. Será feita uma tentativa de como fazer as máquinas usarem a linguagem, formar abstrações e conceitos, resolver tipos de problemas agora reservados aos humanos e melhorar a si mesmas. Acreditamos que um avanço significativo pode ser feito em um ou mais desses problemas se um grupo cuidadosamente selecionado de cientistas trabalhar juntos durante um verão (MCCARTHY *et al.*, 1955, p. 02, tradução nossa).

Durante a conferência, foi apresentado o Logic Theorist, o primeiro programa de computador de IA. Desenvolvido na Corporação de Pesquisa e Desenvolvimento (RAND) por Allen Newell, Cliff Shaw e Herbert Simon, o programa tinha como foco a resolução de problemas matemáticos a partir do livro "*Principia Mathematica*". Desenvolver o Logic Theorist foi um grande desafio para Newell, Shaw e Simon, uma vez que tiveram que trabalhar com o IBM 701, que operava por meio de linguagem de máquina. Para contornar esse obstáculo, eles criaram a linguagem de alto nível IPL (Information Processing Language), que acabou se tornando a linguagem de escolha para muitos projetos de IA por muitos anos. Ademais, a memória limitada do IBM 701 também foi um problema para o Logic Theorist, o que os levou a criar outra inovação: o processamento de listas, que permitiu a alocação e desalocação dinâmica de memória durante a execução do programa (TAULLI, 2019).

Herbert Simon, ao observar o funcionamento dos computadores na impressão de palavras em um mapa para sistemas de defesa aérea, notou que as máquinas tinham o potencial para uma aplicação mais ampla do que simplesmente processar números. Essa percepção foi crucial, pois indicou que essas máquinas poderiam ser empregadas no tratamento de não apenas números, mas também imagens, caracteres e símbolos. Servindo como o alicerce inicial para o desenvolvimento do que hoje conhecemos como IA simbólica, impulsionando avanços notáveis em áreas

como processamento de linguagem natural, visão computacional e resolução de problemas complexos (TAULLI, 2019).

Embora a Conferência de McCarthy em 1956 seja amplamente considerada como o marco zero da IA, as ideias relacionadas a essa área datam de antes de 1956, remontando à Segunda Guerra Mundial. A primeira literatura relevante sobre o tema da IA foi escrita em 1943 por Warren McCulloch e Walter Pitts, que desenvolveram um modelo matemático de estruturas de raciocínio artificiais que imitam o sistema nervoso humano. Esse modelo matemático serviu como base para várias outras formulações acadêmicas sobre o assunto (BARBOSA; BEZERRA, 2020).

O campo da IA tem como seu patrono Alan Turing, um matemático e cientista da computação que em 1950 publicou um estudo intitulado "Máquinas de Computação e Inteligência" (*Computing machinery and intelligence*), onde introduziu os conceitos fundamentais da IA ao abordar a questão: "As máquinas podem pensar?" (TURING, 1950, p. 433, tradução nossa). A partir desse trabalho seminal, Turing propôs o teste de Turing como uma forma de avaliar a capacidade de uma máquina em exibir comportamento inteligente que seja indistinguível do comportamento humano. O teste denominado "Jogo da imitação", consistia em uma máquina capaz de imitar a comunicação escrita de um ser humano. O objetivo do teste era avaliar se a máquina poderia transmitir informações de maneira que se parecesse com uma pessoa real, sem que o receptor suspeitasse que estivesse interagindo com um programa de computador. Turing afirmou que se pelo menos um terço dos participantes fosse enganado e acreditasse que estava dialogando com um ser humano, a máquina poderia ser considerada "inteligente" (TURING, 1950). O estudo de Turing influenciou significativamente o desenvolvimento da IA, tornando-se uma referência fundamental na área (RUSSEL; NORVIG, 2013).

Em 1951, Marvin Minsky desenvolveu a Stochastic Neural Analog Reinforcement Calculator (SNARC), a primeira máquina de rede neural artificial já criada, utilizando componentes analógicos e eletromecânicos e operando com o sistema operacional DOS. A rede consistia de 40 neurônios conectados, projetados individualmente com um capacitor para memória de curto prazo e um potenciômetro para memória de longo prazo. Para avaliar as habilidades de aprendizado da

máquina, Minsky realizou experimentos de navegação em um labirinto virtual, a fim de verificar sua capacidade de aprendizagem e resolução de problemas. A estrutura criada por Minsky serviu como base para a criação de outras estruturas similares em trabalhos futuros (BARBOSA; BEZERRA, 2020).

4.1 PRIMEIRA ONDA DA IA: A ERA DE OURO (1956–1974)

Entre 1956 e 1974, a IA se destacou como um dos temas mais importantes na indústria tecnológica global, sendo reconhecida como "A era de ouro da IA". Esse progresso significativo foi impulsionado em grande parte pela rápida evolução das tecnologias computacionais. Os sistemas de grande porte baseados em tubos de vácuo foram substituídos por sistemas menores, porém mais ágeis, equipados com circuitos integrados de alto desempenho e maior capacidade de armazenamento. Além disso, o governo federal investiu significativamente em inovações tecnológicas. Isso se deve, em parte, aos objetivos ambiciosos do programa espacial Apollo e às intensas demandas da Guerra Fria entre os Estados Unidos e a União Soviética (TAULLI, 2019).

Naquela época, a Agência de Projetos de Pesquisa Avançada (ARPA) era a principal fonte de financiamento para pesquisas em IA. A ARPA foi criada no final dos anos 1950 como resposta ao lançamento do Sputnik pela Rússia, e seus projetos financiados tinham poucas restrições, visando fomentar a inovação revolucionária. J.C.R. Licklider, um dos líderes da ARPA, defendia que o financiamento deveria ser destinado às pessoas, não apenas aos projetos. As universidades de Stanford, MIT, Lincoln Laboratories e Carnegie Mellon foram as principais instituições beneficiadas com o financiamento da ARPA (TAULLI, 2019).

Durante a década de 1950, a IBM direcionou seus esforços para a comercialização de seus computadores, deixando a pesquisa em IA de lado. Isso ocorreu devido ao receio de que a tecnologia resultasse em perda significativa de empregos, evitando assim a responsabilização por esse possível impacto. Como consequência, grande parte da inovação em IA durante esse período foi conduzida por pesquisadores acadêmicos. Nessa época, houve um grande aumento na

produção de artigos e livros acadêmicos sobre IA. (TAULLI; 2019). Entre as principais inovações estão:

- Frank Rosenblatt introduziu em 1957 o *Perceptron*, um algoritmo que consiste em uma rede neural de camada única capaz de classificar resultados (BARBOSA; BEZERRA, 2020);
- Em 1958, John McCarthy desenvolveu a linguagem de programação Lisp, que se tornou uma referência em sistemas de IA na época. Atualmente, a linguagem inspirou uma série de outras linguagens de programação (BARBOSA; BEZERRA, 2020);
- Em 1959, Newell, Shaw e Simon desenvolveram o programa "*General Problem Solver*" que se dedicou a resolver problemas matemáticos, como o quebra-cabeça Torre de Hanói (TAULLI, 2019);
- Em 1959, Arthur L. Samuel trabalhou na IBM, contribuindo para o desenvolvimento do primeiro sistema de computador comercializado pela empresa, o 701, enquanto buscava melhorar a capacidade de processamento das máquinas. Durante esse período, ele também criou um jogo de damas no computador, que se tornou um marco histórico. Esse jogo pioneiro no campo do aprendizado de máquina demonstrou que um computador pode aprender e aprimorar seu desempenho ao processar dados, sem a necessidade de programação específica para cada tarefa. Essa abordagem revolucionária foi viabilizada pela aplicação de conceitos estatísticos avançados, como a análise de probabilidade (TAULLI, 2019);
- Em 1961, o pesquisador do MIT James Slagle desenvolveu o SAINT (Symbolic Automatic INTEgrator), um programa que auxiliava os estudantes calouros na resolução de problemas de cálculo. Posteriormente, esse programa evoluiu para outras versões mais avançadas, conhecidas como SIN e MACSYMA, capazes de realizar cálculos muito mais complexos. Na verdade, o SAINT pode ser considerado o primeiro sistema especialista criado (TAULLI, 2019);

- O professor do MIT Thomas Evans desenvolveu o programa ANALOGY em 1963, que provou que um computador poderia resolver problemas de analogia presentes em testes de QI (TAULLI; 2019);
- Durante sua pesquisa de doutorado sob a supervisão de Minsky no MIT, Daniel Bobrow desenvolveu o programa ESTUDANTE em 1964. Esse aplicativo de IA utilizou técnicas de Processamento de Linguagem Natural (NLP) para resolver problemas de álgebra destinados a estudantes do ensino médio (TAULLI; 2019);
- O professor do MIT, Joseph Weizenbaum, criou o programa ELIZA em 1965, que rapidamente se tornou um sucesso e chamou a atenção da grande imprensa. ELIZA foi nomeado em homenagem à personagem da peça Pygmalion, de George Bernard Shaw, e atuou como uma espécie de psicanalista virtual. Os usuários podiam digitar perguntas e ELIZA forneceria respostas (sendo este o primeiro exemplo de um *chatbot*). Algumas pessoas que usaram o programa acharam que estavam interagindo com uma pessoa real, o que preocupou Weizenbaum, uma vez que a tecnologia subjacente era bastante simples (TAULLI; 2019);
- Marvin Minsky, do MIT, desafiou Gerald Jay Sussman, um de seus alunos, a passar o verão de 1966 conectando uma câmera a um computador e fazendo com que o computador descrevesse o que via. Sussman cumpriu o desafio e criou um sistema denominado *Computer Vision* que detectava padrões básicos, tornando-se assim o primeiro uso conhecido de visão computacional (TAULLI; 2019);
- O programa Mac Hack foi desenvolvido em 1968 pelo professor do MIT Richard D. Greenblatt, foi o pioneiro em jogar xadrez em torneios reais, conseguindo obter uma classificação C (TAULLI; 2019);
- Na década de 1960, o professor Raj Reddy criou o *Hearsay I*, um sistema de reconhecimento de fala contínuo. Alguns de seus alunos fundaram a Dragon Systems, que posteriormente se tornou uma grande empresa de tecnologia na área de reconhecimento de fala (TAULLI; 2019).

Neste período, no campo da IA, surgiram duas teorias principais: uma proposta por Minsky (IA Simbólica), que defendeu a utilização de sistemas simbólicos baseados na lógica computacional tradicional ou em pré-programação, como o uso de declarações *If-Then-Else*; e outra proposta por Frank Rosenblatt (IA Conexionista), que defendeu a utilização de sistemas semelhantes ao cérebro humano, como as redes neurais ou o conexionismo, em que os neurônios são referidos como perceptrons. Nesse sistema, a máquina é capaz de aprender com o tempo ao ser exposta a dados. A abordagem de Rosenblatt se baseia no conceito de que o aprendizado pode ser entendido como o ajuste dos pesos sinápticos entre os neurônios da rede neural (TAULLI; 2019).

Rosenblatt propôs que o aprendizado na IA pode ser alcançado pelo ajuste dos pesos sinápticos entre os neurônios da rede neural, permitindo que a máquina aprenda a reconhecer padrões e executar tarefas complexas à medida que é exposta a dados ao longo do tempo. Essa abordagem é baseada no funcionamento do cérebro humano, em que as conexões entre os neurônios são adaptadas com base na experiência (TAULLI; 2019).

Em 1957, Rosenblatt desenvolveu o primeiro software para IA, denominado Perceptron Mark 1. Esse programa incluía câmeras para distinguir entre duas imagens de 20×20 pixels e seguia um processo em que os dados eram ponderados aleatoriamente e passavam por uma série de etapas até que os resultados fossem precisos. Embora tenha sido um avanço significativo na época, a rede neural do Perceptron possuía apenas uma camada devido à limitação de poder computacional. Além disso, a pesquisa sobre o cérebro e a compreensão da capacidade cognitiva ainda estavam em estágio inicial (TAULLI; 2019).

Minsky e Seymour Papert co-escreveram o livro *Perceptrons* (1969), no qual criticaram duramente a abordagem de Rosenblatt, levando a um rápido declínio do interesse em redes neurais. Minsky havia até criado uma rede neural rudimentar no início dos anos 1950, mas percebeu que a tecnologia ainda estava longe de ser viável. Apesar de Rosenblatt ter tentado refutar as críticas, a comunidade de IA perdeu o interesse em redes neurais. Infelizmente, Rosenblatt faleceu em um acidente de barco alguns anos depois. No entanto, suas ideias foram revividas na

década de 1980, o que levou à revolução na IA com o desenvolvimento do aprendizado profundo (TAULLI; 2019).

Durante a Era de Ouro da IA, houve um ambiente de entusiasmo e liberdade sem precedentes. Acadêmicos renomados em todo o mundo se dedicaram a criar máquinas que pudessem pensar de forma autônoma e resolver problemas complexos com rapidez e precisão. Essa era foi caracterizada por uma série de avanços notáveis, que incluíram a criação de algoritmos de aprendizado de máquina, redes neurais artificiais e sistemas especializados (TAULLI; 2019).

Durante esse período, a IA passou por uma série de revoluções e transformações, com novas abordagens e metodologias sendo desenvolvidas continuamente. Os pesquisadores estavam constantemente experimentando e buscando novas maneiras de criar máquinas mais inteligentes e eficazes, impulsionando o avanço da tecnologia da IA a cada dia (TAULLI; 2019).

No entanto, apesar de todo o entusiasmo e liberdade que caracterizaram a Era de Ouro da IA, os pesquisadores também enfrentaram vários desafios e obstáculos significativos. Lamentavelmente, a próxima etapa da IA se mostraria muito mais obscura. Uma quantidade maior de acadêmicos estava começando a expressar ceticismo somadas a promessas exageradas e expectativas frustradas (TAULLI; 2019).

4.1.1 Primeiro inverno da IA: o descrédito da IA (1974–1980)

Um padrão que se tornou recorrente na comunidade de IA, são ciclos de aproximadamente cinco a dez anos, em que o campo experimenta períodos de grande otimismo - a "Primavera da IA" - seguidos de períodos de escassez de recursos e apoio financeiro - o "Inverno da IA". O ciclo de expectativas elevadas em relação aos avanços da IA seguido de resultados pouco satisfatórios desaceleram o financiamento e pesquisa na área, estabelecendo um ciclo repetitivo de euforia e decepção (MITCHELL, 2019). Os principais invernos ocorreram entre os anos de 1970–1980 e 1987–1993 (RUSSEL; NORVIG, 2013).

O exagero é um fenômeno comum em tecnologias emergentes, que pode levar a uma histeria coletiva (*hype*) em torno de uma nova tecnologia, resultando em promessas exageradas por parte dos desenvolvedores, expectativas irrealistas dos usuários finais e ampla promoção na mídia (NEWQUIST, 2020). A IA passou por alguns períodos de queda temporária de sua reputação durante os chamados "inverno da IA", que foram causados por esses exageros. No entanto, apesar dessa queda na reputação, a IA continuou a se desenvolver com sucesso e a produzir novas inovações. O pesquisador de IA Rodney Brooks argumentou em 2002 que existe um mito equivocados de que a IA falhou nesses períodos, quando na verdade ela está presente em nossas vidas cotidianas (KURZWEIL, 2005).

Existe esse mito estúpido por aí de que a I.A. falhou, mas A.I. está em toda parte ao seu redor a cada segundo do dia. As pessoas simplesmente não percebem isso. Você tem I.A. sistemas em carros, ajustando os parâmetros dos sistemas de injeção de combustível. Quando você pousa em um avião, seu portão é escolhido por um I.A. sistema de agendamento. Toda vez que você usa um software da Microsoft, você tem uma I.A. sistema tentando descobrir o que você está fazendo, como escrever uma carta, e faz um ótimo trabalho. Toda vez que você vê um filme com personagens gerados por computador, eles são todos pequenos I.A. personagens se comportando como um grupo. Toda vez que você joga um videogame, está jogando contra um I.A. sistema (KURZWEIL, 2005, p. 228, tradução nossa).

Ray Kurzweil (2005), enfatiza que muitos erroneamente acreditaram que o inverno da IA poderia ter sido o fim da história, quando, na verdade, há milhares de aplicativos de IA profundamente integrados em todos os setores da infraestrutura atual.

Ainda hoje, muitos milhares de aplicativos de IA estão profundamente incorporados na infraestrutura de todos os setores. A maioria dessas aplicações eram projetos de pesquisa de dez a quinze anos atrás. (KURZWEIL, 2005, p. 229, tradução nossa).

O primeiro inverno da IA ocorreu durante o início dos anos 1970 e se estendeu até meados de 1980. Durante o período, os defensores da IA simbólica

elaboraram diversas propostas de subsídios visando avanços em áreas como compreensão de fala e linguagem, raciocínio de senso comum, navegação de robôs e veículos autônomos (MITCHELL, 2019) (TAULLI, 2019).

Contudo, os relatórios encomendados por agências financiadoras, como o Conselho de Pesquisa Científica do Reino Unido e o Departamento de Defesa dos Estados Unidos, apresentaram resultados desfavoráveis quanto ao progresso e às perspectivas da pesquisa em IA. Embora alguns sistemas tenham sido implementados com sucesso na década de 1970, os avanços mais significativos e promissores na área de IA não foram alcançados conforme o esperado (MITCHELL, 2019) (TAULLI, 2019).

O desenvolvimento da IA neste período teve alguns avanços, no entanto, esses progressos foram principalmente de natureza acadêmica e estavam limitados a ambientes controlados (TAULLI; 2019). Isso significava que, embora algumas melhorias na compreensão dos problemas tivessem sido observadas, ainda não era possível implementar soluções práticas e efetivas em um cenário do mundo real devido algumas limitações tecnológicas, econômicas e sociais.

As limitações dos sistemas computacionais representavam um grande obstáculo para o desenvolvimento da IA na época. A linguagem de programação Lisp, amplamente utilizada na área, não era a escolha ideal para muitos sistemas de computador da época. Como resultado, empresas e organizações estavam predominantemente focadas na linguagem de programação FORTRAN, que era mais adequada para sistemas de computador em geral. Infelizmente, essa dependência limitou ainda mais o progresso em IA, uma vez que as empresas estavam mais preocupadas com a eficiência e desempenho de suas aplicações de negócios do que com o desenvolvimento científico da IA (TAULLI; 2019).

Durante a década de 1970, o ambiente econômico dos Estados Unidos enfrentava vários desafios, como inflação, crescimento lento e interrupções no abastecimento, como a crise do petróleo. Como resultado, o governo dos EUA tomou medidas rigorosas no financiamento. Para um planejador do Pentágono, habilidades como jogar xadrez, resolver teoremas ou reconhecer imagens básicas não eram consideradas tão valiosas quanto outras habilidades. Um exemplo desse

ponto de vista é o Speech Understanding Research desenvolvido pela Carnegie Mellon University. Embora a agência DARPA, responsável por projetos avançados de pesquisa em defesa, tenha acreditado que esse sistema de reconhecimento de fala poderia ser usado por pilotos de caça para emitir comandos de voz, acabou se mostrando impraticável. Um dos programas, chamado Harpy, compreendia apenas 1.011 palavras, o que equivale ao vocabulário de uma criança de três anos. Como resultado, os funcionários da DARPA acreditam que foram enganados e cancelaram o orçamento anual do programa (TAULLI, 2019).

No entanto, a pior repercussão para a IA na época veio através de um relatório, publicado em 1973 pelo professor James Lighthill e financiado pelo Parlamento britânico. O relatório rejeitou completamente os "grandes objetivos" da IA forte, destacando um problema importante conhecido como "explosão combinatória", que tornava os modelos extremamente complicados e difíceis de ajustar. O relatório concluiu que, até então, nenhuma descoberta no campo da IA havia produzido o grande impacto que havia sido prometido anteriormente. "Em nenhuma parte do campo as descobertas feitas até agora produziram o grande impacto que então se prometia." (LIGHTHILL, 1972, p. 9, tradução nossa). O tom pessimista do relatório também questionou a capacidade dos computadores para reconhecer imagens ou vencer um grande mestre de xadrez. O relatório desencadeou um debate público televisionado na BCC, no qual Lighthill apresentou seus argumentos contra Donald Michie, Richard Gregory e John McCarthy. Embora Lighthill tenha apresentado pontos válidos e revisado grandes quantidades de pesquisa, ele não reconheceu o potencial da IA fraca (TAULLI, 2019).

Essas limitações tiveram um impacto direto na capacidade de desenvolvimento da IA na época, em que pesquisadores e empresas enfrentaram dificuldades para obter os recursos e o suporte adequados para seus projetos em IA. Como resultado do inverno da IA, muitos pesquisadores mudaram de carreira, e os que ainda estudavam IA frequentemente se referiam ao seu trabalho com outros termos, como aprendizado de máquina, reconhecimento de padrões e informática (TAULLI, 2019).

4.1.2 Primeira primavera da IA: os sistemas especialistas (1980–1987)

Apesar do inverno da IA, houve importantes avanços tecnológicos, como a criação da retropropagação que é crucial para a atribuição de pesos nas redes neurais, bem como o desenvolvimento da rede neural recorrente (RNN), que possibilita conexões entre camadas de entrada e saída. Além disso, durante as décadas de 1980 e 1990, os sistemas especialistas surgiram como um fator-chave para o avanço da IA na época, impulsionados pelo crescimento explosivo dos PCs e minicomputadores (TAULLI, 2019).

Os sistemas especialistas foram desenvolvidos com base nos conceitos da lógica simbólica de Minsky em uma abordagem conhecida como "conhecimento baseado em regras" ou "sistemas baseados em conhecimento". Eles consistiam em um conjunto de regras e conhecimentos que foram programados por especialistas de domínio para solucionar problemas específicos em suas áreas de atuação. Esses sistemas foram projetados para imitar o raciocínio humano em uma área específica e, portanto, foram capazes de realizar tarefas complexas que exigiam conhecimento especializado (TAULLI, 2019).

O programa DENDRAL de 1967, foi um marco no uso da IA para inferir a estrutura molecular com base nas informações obtidas a partir de um espectrômetro de massa. Ele foi desenvolvido por um grupo de pesquisadores da Universidade de Stanford, liderado por Ed Feigenbaum, Bruce Buchanan e Joshua Lederberg. O programa recebe como entrada a fórmula elementar da molécula e o espectro de massa, utilizando um grande número de regras específicas para prever qual seria o espectro de massa observado para cada possível estrutura molecular. O DENDRAL foi considerado o primeiro sistema especialista bem-sucedido, pois sua habilidade derivou de um grande número de regras de propósito específico obtidas por químicos especialistas. Além disso, o programa foi eficiente porque todo o conhecimento teórico relevante para resolver esses problemas foi mapeado de sua forma geral no componente de previsão de espectro para formas especiais eficientes. A abordagem do DENDRAL foi um exemplo da separação clara entre o conhecimento (na forma de regras) e o componente de raciocínio, que se tornou um tema principal em sistemas posteriores (RUSSEL; NORVIG, 2013).

Outro avanço importante ocorreu em 1972, Feigenbaum, Buchanan e o Dr. Edward Shortliffe desenvolveram o MYCIN, um sistema especialista que tinha como objetivo diagnosticar infecções sanguíneas. Composto por cerca de 450 regras, sua performance foi comparável à de alguns especialistas e superior à de médicos iniciantes. Diferentemente do DENDRAL, o MYCIN não tinha um modelo teórico geral para deduzir as regras. Elas foram obtidas através de entrevistas com especialistas, que adquiriram seus conhecimentos de livros, outros especialistas e da experiência em casos. Além disso, as regras do MYCIN refletiam a incerteza associada ao conhecimento médico, já que esse conhecimento é muitas vezes subjetivo e variável. Dessa forma, o sistema considerava a incerteza na tomada de decisões, o que foi um avanço significativo no desenvolvimento de sistemas especialistas (RUSSEL; NORVIG, 2013).

O Projeto de Programação Heurística (HPP) foi um importante marco na história do desenvolvimento de sistemas especialistas. Feigenbaum e sua equipe de pesquisadores da Universidade de Stanford estavam cientes do enorme potencial que a IA tinha para revolucionar diversos campos do conhecimento humano. Com isso em mente, eles criaram o HPP com o objetivo de investigar a aplicabilidade dessa nova metodologia para além do campo da química, no qual o DENDRAL e o MYCIN foram desenvolvidos (RUSSEL; NORVIG, 2013).

Uma consequência do significativo aumento das aplicações para solução de problemas reais foi o aumento da necessidade por esquemas de representação do conhecimento que fossem práticos e eficientes. Como resultado, houve uma ampla variedade de linguagens de representação e raciocínio que foram desenvolvidas, sendo que algumas dessas linguagens se baseavam em lógica. Por exemplo, a linguagem Prolog ganhou popularidade na Europa, enquanto que nos Estados Unidos a família PLANNER foi bastante utilizada. Algumas linguagens de representação e raciocínio adotaram uma abordagem estruturada, semelhante à taxonomia biológica, em que fatos sobre objetos e eventos específicos eram organizados em uma grande hierarquia. Essa ideia foi inspirada pelos "frames" propostos por Minsky em 1975 (RUSSEL; NORVIG, 2013).

Como relatado, os sistemas especialistas existem desde meados da década de 1960, mas seu uso comercial só se tornou viável na década de 1980. Um

exemplo significativo é o XCON (eXpert CONfigurer), criado por John McDermott na Universidade Carnegie Mellon. O sistema foi projetado para otimizar a seleção de componentes de computador e continha cerca de 2.500 regras, assemelhando-se a um motor de recomendação. Desde o seu lançamento em 1980, o XCON provou ser uma economia significativa de custos para a DEC (Digital Equipment Corporation), pioneira na indústria de computadores, na sua linha de computadores VAX, gerando economias de cerca de \$ 40 milhões em 1986 (TAULLI, 2019).

Após o sucesso do XCON, houve uma explosão na popularidade dos sistemas especialistas, que se transformaram em uma indústria multibilionária. (TAULLI, 2019). O grupo de IA da Digital Equipment Corporation (DEC) entregou 40 sistemas especialistas em 1988 e continuou a produzir outros. A Dupont, por sua vez, tinha em uso 100 desses sistemas e estava desenvolvendo outros 500, com uma economia anual de cerca de 10 milhões de dólares. Naquela época, praticamente todas as principais corporações dos Estados Unidos possuíam um grupo dedicado à IA e estavam explorando o uso ou a pesquisa de sistemas especialistas (RUSSEL; NORVIG, 2013).

Em 1981, os japoneses divulgaram o "Projeto Quinta Geração", um plano de 10 anos para criar computadores inteligentes que utilizassem a linguagem de programação Prolog. Em resposta, os Estados Unidos criaram a Microelectronics and Computer Technology Corporation (MCC), um consórcio de pesquisa projetado para garantir a competitividade nacional em áreas como o projeto de chips e a pesquisa de interfaces humanas. A IA foi incluída como parte de um esforço mais amplo. Na Inglaterra, o relatório Alvey restaurou o subsídio que havia sido cortado após o relatório Lighthill. Apesar dos esforços em todos esses países, os projetos não conseguiram atingir seus objetivos ambiciosos (MCCORDUCK, 2004) (RUSSEL; NORVIG, 2013).

A criação do Deep Blue pela IBM em 1989 foi um marco na história da IA. O objetivo da empresa era desenvolver um computador que pudesse vencer um campeão mundial de xadrez, uma tarefa que parecia impossível para a tecnologia da época. Para alcançar esse objetivo, a IBM utilizou um sistema especialista, que é uma técnica de IA baseada em regras (AMORIM; 2002) (MCCORDUCK, 2004).

O sistema especialista do Deep Blue era capaz de avaliar rapidamente as posições do tabuleiro e determinar qual movimento seria o mais vantajoso. A partir dessas avaliações, o sistema era capaz de prever com precisão os movimentos futuros do oponente. Esse nível de análise era essencial para vencer um jogador como Garry Kasparov, que era conhecido por sua habilidade estratégica e capacidade de antecipar as jogadas do adversário (AMORIM; 2002) (MCCORDUCK, 2004).

O desenvolvimento do Deep Blue foi um projeto ambicioso que levou mais de uma década para ser concluído. Durante esse tempo, a IBM empregou uma equipe de especialistas em IA, matemática e xadrez para trabalhar no projeto. O resultado final foi um computador que processava 200 milhões de posições por segundo, um feito impressionante para a época (AMORIM; 2002).

A vitória do Deep Blue sobre Kasparov em 1996 foi um momento histórico para a IA. Embora tenha havido muita controvérsia sobre a partida, não há dúvida de que o Deep Blue demonstrou a capacidade da tecnologia de vencer um campeão mundial de xadrez. Além disso, o desenvolvimento do Deep Blue e do sistema especialista que o impulsionou pavimentou o caminho para avanços futuros na IA (AMORIM; 2002) (MCCORDUCK, 2004).

De forma geral, a indústria da IA cresceu exponencialmente entre 1980 e 1988, saindo de um mercado avaliado em alguns milhões de dólares para um bilionário. Dentre as centenas de empresas que atuavam nesse ramo, havia aquelas dedicadas à construção de sistemas especialistas, sistemas de visão, robôs, além de softwares e hardwares especializados para tais finalidades (MCCORDUCK, 2004) (RUSSEL; NORVIG, 2013).

4.1.3 Segundo inverno da IA: a queda dos sistemas especialistas (1987–1993)

A década de 1980 e início dos anos 1990 testemunharam um entusiasmo considerável em relação à IA, com várias empresas buscando capitalizar a promessa dessa tecnologia inovadora. No entanto, esse período também trouxe

consigo uma série de desafios que contribuíram para o declínio dessas empresas especializadas em IA (NEWQUIST, 2020).

Um dos principais desafios estava relacionado aos sistemas especialistas, que, embora tenham sido úteis em muitos casos, tinham limitações significativas em termos de aplicação em diferentes áreas. À medida que esses sistemas se tornavam mais complexos, gerenciá-los e atualizá-los com precisão tornou-se uma tarefa complicada, resultando em resultados imprecisos. Além disso, testar esses sistemas se tornou cada vez mais complexo, especialmente quando especialistas de diferentes áreas tinham opiniões divergentes sobre questões importantes. Adicionalmente, esses sistemas não eram capazes de aprender com o tempo, tornando necessárias atualizações constantes nos modelos lógicos subjacentes para manter sua eficácia. Isso levou a um aumento significativo nos custos e na complexidade envolvidos no desenvolvimento e manutenção de sistemas especialistas (TAULLI, 2019).

No final da década de 1980, a popularidade dos sistemas especialistas no mundo dos negócios começou a declinar, o que resultou na fusão ou falência de muitas startups. Esse declínio contribuiu para outro período de estagnação da IA, que durou até cerca de 1993. Com o rápido crescimento do mercado de PCs, houve uma diminuição significativa nas máquinas baseadas em Lisp que dominavam a indústria de hardware de ponta. Além disso, o investimento do governo em IA, como o da DARPA, também diminuiu, em grande parte devido ao fim da Guerra Fria com a queda da União Soviética (TAULLI, 2019).

Os sistemas especialistas foram uma das áreas de pesquisa em IA mais proeminentes durante as décadas de 1980 e 1990, impulsionados pelo crescimento dos PCs e minicomputadores que tornaram o desenvolvimento desses sistemas mais fácil e acessível. No entanto, eles foram criticados por sua falta de flexibilidade e adaptabilidade, uma vez que estavam limitados ao conhecimento que podia ser programado pelos especialistas humanos e também a manutenção e atualização eram difíceis, exigindo a reprogramação manual para refletir mudanças no conhecimento especializado. (RUSSEL; NORVIG, 2013).

Mesmo assim, a abordagem de sistemas especialistas influenciou outras áreas de pesquisa em IA, incluindo a aprendizagem de máquina e a IA distribuída⁵, e continua a ser usada hoje em várias aplicações, como diagnósticos médicos, previsão financeira e automação industrial (RUSSEL; NORVIG, 2013).

Um dos principais desafios enfrentados pelas empresas de IA foi a falta de avanços significativos na tecnologia que pudessem atender às altas expectativas criadas. Muitos dos sistemas de IA da época eram limitados em sua capacidade de aprendizado e não atendiam às demandas do mercado. Além disso, os custos associados ao desenvolvimento e manutenção dessas tecnologias eram frequentemente insustentáveis para muitas empresas (NEWQUIST, 2020).

Outro fator que contribuiu para o declínio das empresas de IA foi a falta de aplicativos comerciais sólidos para a tecnologia. Muitas organizações estavam lutando para encontrar maneiras de incorporar a IA em seus produtos e processos de negócios de uma forma que justificasse os investimentos significativos necessários (NEWQUIST, 2020).

Além disso, as expectativas exageradas em relação à IA, alimentadas pela mídia e por projeções otimistas, levaram a uma bolha especulativa no mercado de IA, o que levou muitas empresas a investir em tecnologias que não estavam prontas para a aplicação prática (NEWQUIST, 2020).

Como resultado desses desafios, muitas das empresas especializadas em IA tiveram dificuldades financeiras, enfrentaram falências ou foram adquiridas por empresas maiores em busca de ativos de tecnologia. Esse período de ajuste e consolidação marcou o fim da primeira fase comercial da IA, caracterizada por um entusiasmo inicial seguido por desafios significativos que ajudaram a moldar a abordagem subsequente para o desenvolvimento e adoção da IA (NEWQUIST, 2020).

⁵ A abordagem distribuída em sistemas de IA visa distribuir o processamento em vários dispositivos interconectados, ao invés de centralizá-lo em um único servidor (SENG; ANG; NGHARAMIKE, 2022). Essa estratégia é particularmente relevante na edge computing, onde o processamento de IA ocorre próximo aos dispositivos de coleta de dados, como sensores e câmeras, em vez de em data centers remotos. Essa abordagem é crucial em aplicações como veículos autônomos, Internet das Coisas (IoT), assistentes virtuais e campos afins (ZHOU et al., 2019).

De acordo com Newquist (2020), até o final de 1993, mais de 300 empresas dedicadas à pesquisa e desenvolvimento em IA haviam enfrentado desafios significativos, incluindo falências, encerramento de operações ou aquisições por outras empresas. Esse período representou um ponto de virada significativo e o fim substancial da primeira onda comercial da IA.

Embora tenha havido uma queda notável no número de empresas exclusivamente dedicadas à IA naquela época, a pesquisa e o desenvolvimento em IA continuaram a avançar em universidades e centros de pesquisa, pavimentando o caminho para o ressurgimento posterior da IA em uma segunda onda comercial, que trouxe avanços tecnológicos significativos e aplicações práticas mais amplas (NEWQUIST, 2020).

4.2 SEGUNDA ONDA DA IA: REDES NEURAIIS (1993–2011)

A partir do início do novo milênio, a indústria de tecnologia alterou o seu foco: em vez de desenvolver tecnologias sofisticadas para a infraestrutura virtual da internet, passou a priorizar a acessibilidade por meio de dispositivos móveis. A busca pela mobilidade tornou-se a principal tendência tecnológica dos anos 2000. No entanto, essa intensa e praticamente universal atenção em smartphones, aplicativos e tecnologia móvel acabou por desviar o foco da discussão sobre IA, que passou a ser cada vez mais abordada apenas em um contexto histórico ou como uma análise retroativa de falhas notáveis (NEWQUIST, 2020).

Apesar de ter sido amplamente ignorada pela indústria de tecnologia, a IA do final da década de 1990 se infiltrou na internet e nas economias móveis. Partes da IA conseguiram sobreviver, principalmente em funções discretas e pouco conhecidas, como corretores ortográficos e gramaticais presentes no Microsoft Word. Os provedores de busca da internet começaram a perceber o potencial da IA para além da pesquisa convencional. Um exemplo é a Alta Vista, que oferecia o Babel Fish, um mecanismo de tradução on-line que permitia aos usuários traduzir páginas da web (desenvolvido como uma última tentativa de IA pela Digital Equipment antes de ser adquirida pela Compaq e, posteriormente, fechada). Outro exemplo é o Navigator, recurso criado pela Netscape que direcionava os usuários

que visitavam uma página para outras com conteúdo semelhante (NEWQUIST, 2020).

Geoffrey Hinton, um pesquisador da Universidade de Toronto, sempre defendeu a capacidade das redes neurais, desenvolvidas por Frank Rosenblatt (IA Conexionista), de aprender como os seres humanos, reconhecendo imagens e padrões, em vez de depender da lógica convencional. Seu objetivo era permitir que as redes neurais identificassem múltiplos padrões por meio de um processo combinado de treinamento inicial, tentativa e erro e uso de processadores não convencionais (NEWQUIST, 2020).

No entanto, em 1969, Marvin Minsky e Seymour Papert (IA Simbólica), desacreditaram as redes neurais em seu livro "Perceptrons", colocando-as no patamar mais baixo da IA. Desde então, a comunidade de pesquisa nunca apoiou as redes neurais tanto quanto apoia os sistemas especialistas. De fato, no início do século XXI, muitos periódicos de associações profissionais não publicavam artigos sobre redes neurais (NEWQUIST, 2020).

Hinton observou que a limitação primária para o progresso da IA consistia na capacidade de processamento dos computadores. No entanto, ele notou que a Lei de Moore⁶ previa um aumento exponencial no número de componentes em um chip a cada 18 meses, o que poderia favorecer o desenvolvimento da IA nos próximos anos (TAULLI, 2019).

Durante várias décadas, Geoffrey Hinton dedicou-se incansavelmente ao estudo das redes neurais, apesar de enfrentar resistência em relação à sua pesquisa. Como forma de superar essa adversidade, ele adotou o termo "aprendizado profundo" (deep learning) para se referir à sua abordagem, que fazia uso das chamadas "redes de crenças profundas". Essas redes consistiam em redes neurais multicamadas com o propósito de reconhecer padrões através do treinamento com dados de entrada específicos, como, por exemplo, a identificação de rostos humanos. Para atingir o nível de precisão desejado, o sistema empregava

⁶ Formulada por Gordon Moore, em 1965, é uma observação que prevê que o número de transistores em um chip de circuito integrado dobrará a cada dois anos, enquanto os custos por transistor diminuirão (MOORE, 1988). Em outras palavras, a capacidade de processamento dos chips aumentaria exponencialmente ao longo do tempo, enquanto os custos de fabricação diminuiriam.

algoritmos que consideravam as falhas e suposições equivocadas da rede, atribuindo pesos diferentes às equações que a ajudaram a chegar às respostas corretas e diminuindo o valor das equações que levavam a respostas incorretas. Em 2006, Hinton e sua equipe publicaram um artigo sobre as "redes de crenças profundas", embora essa técnica não tenha sido diretamente rotulada como uma rede neural multicamada (NEWQUIST, 2020).

A pesquisa inovadora de Hinton, que se baseou nas realizações de outros pesquisadores entusiastas de redes neurais, foi responsável por desencadear uma série de importantes avanços. Segundo Taulli (2019), esses avanços incluem:

- Em 1980, Kunihiko Fukushima desenvolveu o Neocognitron, um sistema de detecção de padrões que serviu como fundamento para as redes neurais convolucionais, cuja inspiração foi obtida do córtex visual dos animais (TAULLI, 2019).
- John Hopfield criou as "Hopfield Networks" em 1982, que são redes neurais recorrentes (TAULLI, 2019).
- Em 1989, Yann LeCun combinou redes neurais convolucionais com retropropagação, sendo que essas foram utilizadas para análise de cheques manuscritos (TAULLI, 2019).
- Christopher Watkins descreveu o Q-Learning em sua tese de doutorado "Aprendendo com recompensas atrasadas" em 1989, o que promoveu um grande avanço no aprendizado por reforço (TAULLI, 2019).
- Yann LeCun publicou "Aprendizado baseado em gradiente aplicado ao reconhecimento de documentos" em 1998, que empregou algoritmos de descida para aprimorar as redes neurais (TAULLI, 2019).

Hinton teve a oportunidade de empregar sua tecnologia na renomada competição ImageNet, que foi iniciada em 2007 por Fei-Fei Li, uma pesquisadora da Universidade de Princeton. O projeto ImageNet apresentava imagens de uma vasta

variedade de itens, desde gatos até cachoeiras, e não se limitava a imagens individuais, mas sim incluía milhares de variações de cada item. Li concebeu o projeto com o intuito de melhorar a capacidade dos programas de visão computacional em reconhecimento visual, por acreditar que quanto mais fontes de entrada um programa e seus algoritmos pudessem acessar, maior seria sua habilidade de reconhecer novas, inesperadas e futuras versões do item em questão (NEWQUIST, 2020).

Li liderou uma equipe que criou um banco de dados de mais de três milhões de imagens usando recursos acessíveis do mercado de crowdsourcing da Amazon Mechanical Turk, nesse contexto, as imagens eram identificadas e marcadas por usuários individualmente por um preço baixo em centavos. Eles realizaram uma marcação e preparação precisa, específica e imparcial dos dados para garantir que pudessem ser manipulados visualmente de forma correta e eficiente pelos computadores. Em 2009, ela escreveu um artigo sobre o potencial do ImageNet para treinar algoritmos de reconhecimento, no entanto, o trabalho não recebeu muita atenção na época. A ideia evoluiu de um projeto de banco de dados para uma competição, onde as imagens poderiam ser utilizadas por qualquer pesquisador para treinar seus próprios programas de visão computacional. Os programas seriam então testados com novas imagens para avaliar a capacidade de identificação de classes de objetos (NEWQUIST, 2020)

A empresa Cycorp, que pertence à primeira onda de empresas de IA, tem um plano semelhante ao da competição ImageNet, mas com um objetivo diferente. Eles estão trabalhando há quase quatro décadas em um banco de dados chamado CYC, que busca conter todo o conhecimento disponível no mundo. A ideia é que sistemas especialistas possam extrair informações desse banco de dados para solucionar diversos problemas. Ao contrário da competição ImageNet, em que os sistemas de reconhecimento determinam por si mesmos o que torna cada imagem única, o sistema CYC já define previamente quais características compõem um determinado objeto (NEWQUIST, 2020).

O presente marco histórico na IA marca a transição da primeira onda para a segunda onda moderna. Na primeira geração, os sistemas inteligentes eram programados com todo o conhecimento e habilidades necessários, juntamente com

heurísticas derivadas de um assunto específico. Isso envolvia o pré-carregamento de informações, permitindo que o sistema especialista buscasse respostas pesquisando toda a experiência codificada. Na nova abordagem, a ideia de fornecer todas as informações e o conhecimento de um especialista foi abandonada, em vista do alto consumo de tempo. Em vez disso, o sistema é treinado com exemplos e aprende por meio de tentativa e erro, distinguindo entre respostas corretas e incorretas. Essa abordagem é conhecida como aprendizado profundo em redes neurais e representa uma mudança significativa do aprendizado supervisionado para o não supervisionado (NEWQUIST, 2020).

De acordo com Newquist (2020), o processo de aprendizado profundo pode ser comparado ao processo de montar um quebra-cabeça, no qual as peças representam dados que devem ser agrupados de maneira a formar uma imagem coerente. O processo envolve separar as peças em grupos e tentar encaixá-las nas formas corretas, em um processo de tentativa e erro que requer várias tentativas até que a imagem final seja formada. Nas redes neurais, o objetivo é a imagem que deve ser formada com base nos pontos de dados não classificados. Esses pontos de dados são divididos em diferentes camadas para identificar padrões e relacionamentos. A rede neural usa tentativa e erro para encontrar a solução, recompensando as partes do algoritmo que resultam em sucesso. Em contraste com a habilidade humana de entender os elementos da imagem e das peças do quebra-cabeça, a rede neural usa matemática para identificar os padrões e relacionamentos.

4.2.1 Segunda primavera da IA: o *big data* e aprendizado profundo (2011–2022)

A participação do computador Watson, criado pela IBM, no programa de televisão Jeopardy! em fevereiro de 2011 foi outro marco significativo para a IA. Watson jogou contra Ken Jennings e Brad Rutter, dois humanos que são considerados os jogadores mais vitoriosos da história do programa. Watson foi capaz de vencer a competição em uma partida de dois dias transmitida nacionalmente, demonstrando sua habilidade em processamento de linguagem natural e amplo conhecimento. A vitória de Watson no Jeopardy! foi vista como um

momento transcendente para muitos entusiastas da IA. Embora Watson tenha cometido alguns erros, ele ainda conseguiu superar seus adversários humanos e ganhar a partida.

A vitória de Watson foi considerada como um marco importante para a IA na época, no entanto, foi em 2015 que a IA ressurgiu definitivamente (NEWQUIST, 2020). Para entender o que tornou 2015 um ponto de inflexão, é necessário identificar os trabalhos e eventos notáveis que foram responsáveis pelo retorno bem-sucedido da IA ao centro das atenções.

Ao longo dos últimos 15 anos, o Google se consolidou como uma das empresas mais importantes no campo da IA. Esse destaque é resultado da necessidade de lidar com o crescimento exponencial da indexação da Web, o que exigiu da empresa o desenvolvimento de métodos criativos para construir sistemas escaláveis. Assim, foram criadas inovações como clusters de servidores de baixo custo, virtualização e software de código aberto. Outro fator que contribuiu para o sucesso do Google em IA foi a adoção pioneira do aprendizado profundo, uma técnica de aprendizado de máquina que permite o treinamento de modelos computacionais capazes de realizar tarefas complexas, como reconhecimento de fala e imagem. Em 2011, a empresa lançou o projeto "Google Brain" para aplicar essa tecnologia em seus produtos e serviços (TAULLI, 2019).

A assistente digital Siri foi desenvolvida pela Apple em 2011 para o iPhone, com a capacidade de reconhecer a fala do usuário, fornecer respostas para perguntas e realizar anotações e ditados. A Siri é considerada a primeira implementação amplamente adotada de reconhecimento de voz para o consumidor. Ao incorporar tecnologia de reconhecimento de voz e a ideia de um assistente inteligente em um dispositivo móvel, a Apple criou um novo conceito de assistente pessoal. É interessante notar que, durante o segundo inverno da IA, o telefone celular ainda não era considerado uma mercadoria de consumo viável. Como resultado, a IA encontrou uma nova oportunidade de vida em aplicativos móveis (NEWQUIST, 2020).

Entre 2012 e 2015, houve uma disseminação global de programas e artigos que mencionavam as redes neurais. Elas foram empregadas em diversas áreas, tais

como o reconhecimento de padrões de fala, a tradução de frases, a seleção de candidatos a emprego, a realização de compras publicitárias, a previsão de vendas, a escolha de ações e o diagnóstico de doenças. Um exemplo notável é o enfoque do Baidu em padrões de fala para aprimorar o reconhecimento de voz, a fim de expandir seus negócios de mecanismos de busca por meio dessa tecnologia, uma vez que digitar caracteres em mandarim em dispositivos móveis é bastante complexo (NEWQUIST, 2020).

Nos anos iniciais da competição ImageNet, os programas de reconhecimento e visão computacional eram comumente programados manualmente. Entretanto, em 2012, a equipe liderada por Hinton adotou uma abordagem diferente ao empregar uma rede neural conhecida como AlexNet para executar tarefas de reconhecimento. Em vez de programar cada etapa do processo, a equipe introduziu seus algoritmos em imagens de treinamento do ImageNet e permitiu que o AlexNet aprendesse, por si só, a identificar as características de cada imagem. Isso resultou na criação de um código exclusivo pela rede neural. Durante o concurso, a aplicação do AlexNet resultou em uma vitória notável, demonstrando uma taxa de reconhecimento superior a 84%, muito acima dos resultados obtidos por quaisquer outros programas utilizados anteriormente. Como referência, a taxa de reconhecimento total por humanos foi de aproximadamente 95%, fornecendo uma comparação significativa. (NEWQUIST, 2020).

Utilizando uma versão aprimorada da rede neural, que havia sido anteriormente ridicularizada, a equipe liderada por Hinton desenvolveu um sistema de reconhecimento de padrões que superou todos os métodos anteriores. Esse feito foi alcançado com o apoio do CIFAR, o Instituto Canadense de Pesquisa Avançada, localizado no Canadá, superando projetos de IA renomados como o MCC e o Projeto de Quinta Geração do Japão. O sucesso de Hinton e suas redes neurais atraiu a atenção da comunidade global de tecnologia (NEWQUIST, 2020).

Então, sem aviso, a inteligência artificial voltou gritando. Pode-se usar qualquer número de analogias, desde a fênix ressurgindo das cinzas até zumbis e mortos-vivos. O fato é que ninguém previu o quão importante a IA se tornaria para o negócio de tecnologia e com que rapidez ela se reinseriria na conversa global – em um nível muito mais alto do que nas décadas de 1980 e 1990 (NEWQUIST, 2020, p. 626-627, tradução nossa).

Em 26 de maio de 2017, o AlphaGo, um programa de IA desenvolvido pela DeepMind, uma divisão da Alphabet, controladora do Google, alcançou uma notável vitória sobre o maior jogador de Go do mundo, Ke Jie, em uma partida histórica. O Go é considerado um jogo altamente complexo, que exige habilidades avançadas em torneios e estratégias mentais. Anteriormente, especialistas haviam afirmado que a IA não seria capaz de superar os melhores jogadores humanos. No entanto, o triunfo do AlphaGo sobre Ke Jie, por meio da mais recente geração de redes neurais, evidenciou a capacidade da IA de superar limites anteriormente intransponíveis (NEWQUIST, 2020).

Embora as redes neurais tenham alcançado uma série impressionante de vitórias em um período curto, não se pode dizer que seu sucesso tenha sido instantâneo, pois exigiu mudanças drásticas nos fundamentos da indústria de computação para permitir seu ressurgimento (NEWQUIST, 2020). Segundo Newquist (2020), o ressurgimento da IA, juntamente com o renascimento das redes neurais, foi resultado da convergência de quatro fatores:

- 1) GPUs baratas, que proporcionaram um poder de processamento mais rápido;
- 2) Redes neurais profundas que tiraram proveito dessas GPUs;
- 3) Enormes conjuntos de dados, como os compilados diariamente pelo Google e Amazon;
- 4) E à disposição de grandes organizações, como Google, Amazon, Apple, IBM e o gigante chinês da web Baidu, em explorar a IA, independentemente de saberem como ela seria comercializada.

A interação harmônica desses elementos foi fundamental para devolver a IA ao topo da hierarquia tecnológica. Cada um deles contribuiu de maneira igual para o avanço da IA e a falta de qualquer um desses elementos teria reduzido drasticamente a probabilidade de a IA atingir a mesma massa crítica (NEWQUIST, 2020).

As GPUs (Unidades de Processamento Gráfico) são dispositivos de hardware projetados originalmente para lidar com aplicativos com uso intensivo de gráficos, como jogos de videogame. Esses programas apresentam muitos segmentos pequenos que requerem atualizações rápidas, mas não são muito complexos individualmente. Para lidar com a grande quantidade de dados gráficos exigidos pelos jogos de videogame, as GPUs são capazes de processar simultaneamente e rapidamente esses pequenos, mas numerosos, elementos e pixels em paralelo. Essa abordagem altamente eficiente também é útil para outras tarefas que exigem processamento paralelo de grandes quantidades de dados, como cálculos, simulações e renderizações complexas (NEWQUIST, 2020).

Historicamente, a programação e o processamento de pesquisa de alto nível eram feitos em hardware dedicado de ponta ou em PCs ampliados equipados com vários processadores. No entanto, a execução desses programas exigia o uso de hardware caro baseado em CPU, e a programação para essas mesmas CPUs exigia software complexo. A partir de 2004, pesquisadores da Microsoft e de Stanford começaram a quebrar seus complexos algoritmos de aprendizado de máquina para que pudessem ser processados de forma barata e rápida usando as GPUs incorporadas aos computadores pessoais, principalmente para acelerar as cenas nos videogames. Embora a codificação para GPUs seja mais desafiadora do que para CPUs, permite dividir a solução de problemas ou o processamento em pequenos bits. A NVIDIA, fabricante de GPUs, reconheceu que muitos programadores que não eram de videogame estavam usando seus produtos e introduziu uma linguagem de programação chamada CUDA para facilitar a programação de seus chips. Com CUDA, os programadores podem escrever programas em linguagem C que aproveitam o poder de processamento paralelo das GPUs da NVIDIA (NEWQUIST, 2020).

Portando, as GPUs (Unidades de Processamento Gráfico) são processadores compostos por vários núcleos dedicados ao processamento de informações simultâneas, enquanto as CPUs (Unidades Centrais de Processamento) são processadores que tradicionalmente operam com um núcleo central de processamento, realizando operações de maneira linear. Newquist (2020) compara as GPUs à uma metralhadora, já que conseguem processar muitas pequenas

operações poderosas em simultâneo, enquanto as CPUs são comparáveis a uma bazuca, sendo capazes de realizar uma única operação complexa de cada vez, necessitando de recarga para novas operações. Ademais, as GPUs tendem a ser mais acessíveis economicamente se comparadas às CPUs (NEWQUIST, 2020).

A união entre GPUs e redes neurais marcou o segundo grande avanço da segunda onda de IA. Ambas as tecnologias são resultado da aplicação do paralelismo. As GPUs ofereceram um nível de paralelismo que era difícil de ser alcançado com semicondutores lineares tradicionais, e as redes neurais abordaram problemas de maneira altamente paralela, o que contrastava com a abordagem de programação utilizada em sistemas especializados (NEWQUIST, 2020).

Uma rede neural é um modelo matemático inspirado no funcionamento do cérebro humano que é utilizado para realizar tarefas de aprendizado de máquina. Essas redes são treinadas com dados em nível granular, ou seja, com informações mais detalhadas e específicas, para que possam aprender a reconhecer padrões e fazer previsões. À medida que as redes neurais aprendem com esses dados, elas constroem camadas ricas em informações para conseguir encontrar soluções cada vez mais precisas. Isso significa que a rede neural processa os dados em várias camadas, cada uma com uma quantidade crescente de informações relevantes. No entanto, a criação de várias camadas em uma rede neural pode ser computacionalmente intensiva, o que pode ser um problema, uma vez que o tempo de treinamento pode levar horas ou até dias. Porém, o uso de GPUs permite que as redes neurais processem esses dados em paralelo e de forma muito mais rápida, permitindo milhões de iterações do ciclo de aprendizado em uma fração do tempo esperado. Assim, as redes neurais são construídas com várias camadas e os dados são processados em níveis granulares para que a rede neural possa aprender e fazer previsões cada vez mais precisas. O uso de GPUs permite que essas redes neurais possam ser treinadas de forma mais eficiente e em menor tempo (NEWQUIST, 2020).

O terceiro avanço fundamental para o ressurgimento da IA é a utilização de conjuntos de dados. Anteriormente, grandes conjuntos de dados relacionados eram raros, e os maiores bancos de dados eram registros corporativos administrativos, estoques e registros básicos de clientes. No entanto, surgiram empresas de

tecnologia, como Google, Amazon, Facebook e YouTube, que começaram a coletar enormes quantidades de dados de usuários. Esses dados continham informações detalhadas sobre as preferências do consumidor, padrões de compra, 'curtidas' e 'seguidas', favoritos de entretenimento e frequência de uso. Os datafarms, que utilizavam milhares de servidores, armazenavam quantidades quase incalculáveis de dados que a internet produzia a cada segundo de cada dia. Os dados foram marcados pelos próprios usuários e serviram como entrada perfeita para treinar redes neurais, que são um tipo de algoritmo de aprendizado de máquina projetado para reconhecer padrões nos dados. Empresas como Google e Amazon utilizaram esses dados para melhorar seus algoritmos de IA, pré-identificando e classificando milhões de rostos, frases, produtos, dentre outros (NEWQUIST, 2020).

Em 2011, o Google Brain empregou redes neurais em seu repositório de milhões de vídeos do YouTube. A amostra de treinamento consistiu em 10 milhões de vídeos do YouTube, não marcados, não rotulados e não identificados, durante três dias. As redes neurais foram alimentadas com 20.000 categorias de objetos reais marcados e identificados como metas. Posteriormente, as redes neurais foram distribuídas em 16.000 processadores sem supervisão para categorizar ou identificar as 10 milhões de imagens do YouTube, com base nos 20.000 objetos marcados, sem ter recebido instruções prévias sobre a relação entre eles (NEWQUIST, 2020).

Ainda que a classificação pareça uma tarefa fácil, algo que até mesmo uma criança pequena poderia realizar, sua dificuldade aumenta significativamente devido ao tamanho e complexidade dos dados. Adicionalmente, a falta de familiaridade prévia ou de um ponto de referência torna a tarefa ainda mais desafiadora (NEWQUIST, 2020). Segundo Newquist (2020, p. 638, tradução nossa), é como se alguém dissesse "aqui estão dez milhões de fotos e você nunca viu os objetos nessas fotos antes. Agora, descubra em qual das vinte mil categorias cada uma dessas fotos se enquadra." A máquina precisa começar do zero, identificando e caracterizando objetos de formas básicas, como objetos redondos, antes de poder classificá-los em categorias mais específicas.

Imagine ter que classificar objetos que você nunca viu antes, como caracteres do alfabeto balinês ou hieróglifos egípcios, e determinar como eles estão relacionados. Essa é uma tarefa assustadora, mesmo para seres humanos muito

inteligentes, pois requer um conhecimento prévio dos objetos e do contexto em que são usados. No entanto, as redes neurais do Google conseguiram superar esse desafio, vasculhando milhões de imagens para identificar objetos e classificá-los corretamente. Elas aprenderam que um item específico em quase qualquer foto pode ser identificado como algo específico. Por exemplo, as redes foram capazes de identificar rostos humanos com uma precisão de 81% e corpos humanos com uma precisão de 76%, mesmo sem saber o que constituía um humano. Elas também foram capazes de identificar gatos com uma precisão de quase 75%. No entanto, é importante notar que as redes neurais não entendem o contexto ou o significado dos objetos que estão classificando. Elas simplesmente identificam características comuns em pixels de imagens e usam a matemática para agrupar objetos visualmente semelhantes em categorias. Por exemplo, a rede neural não sabe que uma imagem é de um gato, mas reconhece características específicas, como formas geométricas, contornos e outras características comuns em imagens de gatos. Com base nessas características, a rede neural agrupa objetos visualmente semelhantes em categorias. Portanto, embora a precisão da rede neural seja impressionante, é importante lembrar que ela não possui um entendimento completo do que está classificando. A rede neural simplesmente faz previsões com base em padrões matemáticos identificados nas imagens que analisa (NEWQUIST, 2020).

A existência de grandes conjuntos de dados (*big data*), especialmente aqueles que possuem informações descritivas, tem sido fundamental para o avanço da tecnologia de IA. Gigantes corporativos como Google, Amazon, Facebook e YouTube têm coletado vastas quantidades de informações dos usuários, as quais são utilizadas para treinar suas redes neurais e melhorar seus algoritmos. Como resultado, a IA tem conseguido progressos significativos no reconhecimento de padrões em dados, na classificação de imagens e objetos e na realização de tarefas que antes eram consideradas exclusivas dos seres humanos. Taulli (2019, p.18, tradução nossa) afirma que os dados são "o combustível para a IA".

O uso de redes neurais para processamento de grandes volumes de dados culminou em sistemas de recomendação acionados por correspondência de IA, que se apresentam como solução prática para o consumidor. As sugestões de recomendação, tais como qual música ouvir, qual vídeo assistir ou qual livro

comprar, são possíveis graças à habilidade da IA em realizar conexões entre grandes quantidades de dados que anteriormente estavam inativos. Pode-se inferir que os dados estavam à espera de serem manipulados por uma máquina inteligente, pois nenhum grau de inteligência humana seria capaz de prever ou manipular os desejos e caprichos de milhões de usuários diários. Em alguns casos, os usuários nem mesmo tinham conhecimento de seus próprios desejos e caprichos até serem notificados pelos sistemas de recomendação (NEWQUIST, 2020).

O investimento e o comprometimento em IA por parte das grandes empresas resultaram no quarto fator. Pois à medida que as redes neurais demonstravam aos gigantes da tecnologia como utilizar dados para maximizar o tempo de navegação, o valor monetário e a retenção de visitantes humanos em suas páginas, eles aumentavam cada vez mais o investimento em pesquisas relacionadas à IA. Como resultado, a divisão Watson da IBM se tornou uma das mais destacadas, o Google lançou o projeto Google Brain, e a Intel e a Microsoft criaram grupos que afirmavam ser o futuro de suas empresas (NEWQUIST, 2020).

Uma série de avanços significativos na área da IA começaram a se tornar cada vez mais presentes e difundidos em nosso cotidiano. Esses feitos da IA passaram a se disseminar rapidamente. Um exemplo mencionado é o serviço de tradução automática de idiomas do Google, conhecido como Google Tradutor. Embora não tenha sido perfeito inicialmente, ele funcionou de maneira surpreendentemente eficaz e, ao longo do tempo, passou por melhorias significativas. Isso possibilitou uma melhor comunicação e compreensão entre pessoas de diferentes idiomas, facilitando a troca de informações e a colaboração global (MITCHELL, 2010).

Outro exemplo é o desenvolvimento dos veículos autônomos do Google, que começaram a circular pelas estradas do norte da Califórnia. Embora tenham sido cautelosos e hesitantes no começo, eles eram capazes de se mover de forma independente no tráfego. Isso representou um avanço importante no campo da IA, abrindo caminho para a futura implementação de veículos autônomos em grande escala (MITCHELL, 2010).

Além disso, assistentes virtuais como a Siri da Apple e a Alexa da Amazon se tornaram amplamente disponíveis em dispositivos móveis e residências. Esses assistentes virtuais podem atender a muitas solicitações por meio de comandos de voz, desde tarefas simples até a obtenção de informações e o controle de dispositivos inteligentes domésticos (MITCHELL, 2010).

Outros avanços incluem o fornecimento de legendas automáticas altamente precisas pelo YouTube, a tradução simultânea entre idiomas durante videochamadas pelo Skype e o aprimoramento do reconhecimento facial do Facebook, capaz de identificar rostos com notável precisão em fotos enviadas. O site de compartilhamento de fotos Flickr também implementou recursos de marcação automática, onde as fotos são identificadas e descritas automaticamente através de texto que descreve seu conteúdo (MITCHELL, 2010).

Todos esses eventos ocorreram sem serem expostos ao público e, em muitos casos, passaram despercebidos pelos próprios funcionários dessas empresas que estavam desenvolvendo sistemas de IA. Foi somente em 2015 que a comunidade de tecnologia, investidores e mídia notaram que a IA era uma força a ser reconhecida. Neste ano, os programas da Google e da Microsoft alcançaram níveis de precisão de reconhecimento acima de 95% na competição ImageNet, provando que haviam desenvolvido sistemas superiores aos humanos na capacidade de reconhecer objetos. Essas conquistas foram possíveis graças ao uso de redes neurais multicamadas, que comprovaram o valor da estratégia inicial de Hinton. Foi uma clara demonstração de que os humanos foram superados em uma habilidade essencial, e não apenas em um jogo (NEWQUIST, 2020).

Nos últimos anos, houve um significativo progresso no avanço da IA, resultando em impactos notáveis na indústria e economia global. Em 2015, muitas startups de IA começaram a mostrar progressos notáveis na habilidade de reconhecer padrões ou fornecer sugestões baseadas nos dados do usuário. No entanto, essas empresas foram rapidamente adquiridas pelos principais *players* da tecnologia, como Google, Apple, Intel, Microsoft, Facebook, Salesforce, Baidu, Amazon e X (Twitter), que passaram a adquirir startups de IA em todo o mundo. Os principais *players* da tecnologia adentraram os laboratórios de pesquisa universitária em um ritmo frenético, liderando a carga nos EUA e na China, seguidos pelo Reino

Unido, Índia, Israel, França e Alemanha. A importância da IA era tamanha que as maiores corporações mundiais não poderiam se dar ao luxo de ficar atrás na pesquisa de IA. Em 2015, as aquisições de empresas de IA ocorriam quase a cada duas semanas nos Estados Unidos. No final de 2017, apenas a Intel já havia investido mais de um bilhão de dólares em IA, desde pesquisa até startups, superando a receita total de todas as empresas de IA da primeira onda - em conjunto. A comunidade de investimentos injetou US\$6 bilhões em startups de IA entre 2014 e 2017. Na primeira onda, US\$6 milhões em investimentos seriam considerados extremamente elevados (NEWQUIST, 2020).

Atualmente, a IA está presente em praticamente todos os setores, desde carros autônomos até traduções de documentos online em mais de cem idiomas. O Google, por exemplo, tem mais de mil projetos relacionados à IA, incluindo a divisão de carros autônomos Waymo. A empresa até mesmo criou seus próprios chips de IA - chamados *Tensor Processing Units* (tensores são objetos matemáticos que podem ser processados para descrever propriedades físicas) - e a Alphabet, empresa controladora do Google, nem é considerada um produtor de hardware (NEWQUIST, 2020).

De acordo com Mitchell (2019), o entusiasmo em torno da IA neste período estava crescendo rapidamente e o mundo dos negócios estava prestando muita atenção nisso. As principais empresas de tecnologia investiram quantias significativas de dinheiro em pesquisas e desenvolvimento de IA, seja contratando especialistas em IA diretamente ou adquirindo pequenas startups para ter acesso aos seus talentosos funcionários. A possibilidade de ser adquirido, o que oferecia a perspectiva de se tornar um milionário instantâneo, estimulou o surgimento de inúmeras startups lideradas por ex-professores universitários, cada uma com sua própria abordagem em relação à IA. Mitchell (2019, p. 46, tradução nossa) complementa: "como observou o jornalista de tecnologia Kevin Kelly, 'os planos de negócios das próximas 10.000 startups são fáceis de prever: pegue X e adicione IA'".

Pesquisadores como Fei Fei Li, Yann LeCun, Andrew Ng, Yoshua Bengio, Michael I. Jordan, Zoubin Ghahramani, Haifeng Wang, Daphne Koller e David Blei se tornaram Minsky, McCarthy, Shannon, Newell e Simon da nova geração. Como seus antecessores de IA, eles foram retirados dos laboratórios e assumiram papéis empresariais. Ao contrário de seus predecessores de IA, a maioria deles era respaldada por bilhões de dólares em receita gerada pelos gigantes da tecnologia. Com todo aquele dinheiro, eles iriam fazer coisas incríveis. E algumas coisas estúpidas também (NEWQUIST, 2020, p. 641 , tradução nossa).

4.2.2 Verão da IA: a IA generativa (2022–presente)

No início dos anos 2020, um período notável de progresso surgiu no campo da IA foi impulsionado principalmente pelo desenvolvimento de redes neurais profundas baseadas em *transformers* (CAO *et al.*, 2023). Esses avanços abriram caminho para uma multiplicidade de sistemas de IA generativa que exibiam uma capacidade diferente das demais: eles podiam compreender e responder a *prompts* (comandos) de linguagem natural como entrada.

O surgimento desses sistemas baseados marcou uma mudança significativa em relação às tecnologias de IA anteriores. Sistemas de IA tradicionais tinham dificuldade em compreender as nuances da linguagem humana e gerar respostas coerentes. No entanto, o advento das arquiteturas de *transformers*, com seus mecanismos de atenção e redes neurais profundas, permitiu que esses sistemas não apenas entendessem a linguagem natural, mas também gerassem texto que era contextualmente relevante, coerente e frequentemente indistinguível de conteúdo escrito por humanos (KAHN *et al.*, 2022).

Os modelos de linguagem baseados na aprendizagem (LLMs), também conhecidos como modelos de processamento de linguagem natural, representam uma classe revolucionária de sistemas de IA. No contexto dos LLMs, essa aprendizagem se concentra na linguagem natural, a maneira como os seres humanos se comunicam e interagem por meio de palavras e textos. Eles são treinados em grandes volumes de texto, como livros, artigos, sites e conversas

humanas, o que lhes permite desenvolver uma compreensão profunda das nuances e complexidades da linguagem humana (CHEN et al., 2023).

Esses modelos de distribuição de probabilidade e processamento de linguagem natural têm sido responsáveis por algumas das conquistas mais notáveis na IA recentemente. Essa mudança de paradigma proporcionou uma nova forma de interação entre humanos e computadores e expressão criativa. Alguns dos sistemas de IA mais notáveis que surgiram durante esse período incluíram o ChatGPT, Bing Chat (Microsoft em parceria com a OpenAI), Bard (Google) e LLaMA (Meta/Facebook), bem como sistemas de arte de inteligência artificial texto-imagem, como Stable Diffusion, Midjourney e DALL-E (MONDAL; DAS; VRANA, 2023).

Um exemplo notável é o ChatGPT, lançado em 2022 pela OpenAI, um renomado centro de pesquisa em Inteligência Artificial, inicialmente estabelecido por Sam Altman, Greg Brockman, Elon Musk e outros, que rapidamente ascendeu à fama como uma poderosa ferramenta de conversação baseada em IA. Sua versatilidade era notável, uma vez que demonstrava a capacidade de participar em diálogos com profundidade e significado (CHATTERJEE; DETHLEFS, 2023).

O LLM utilizado no ChatGPT, conhecido como GPT-3 (*Generative Pre-trained Transformer - 3*), é uma rede neural com cerca de 175 bilhões de parâmetros e foi treinado com uma vasta quantidade de dados da *internet*, que somam mais de 75 terabytes (CHATTERJEE; DETHLEFS, 2023). Com essa capacidade, o modelo é capaz de gerar textos e respostas de maneira autônoma, sem precisar de muitos exemplos específicos para cada tarefa.

Segundo Wolfram (2023), o principal propósito do ChatGPT (e das maiorias IA generativas) é o de gerar de forma consistente uma continuação textual que esteja em perfeita consonância com o contexto já estabelecido. Quando se fala em "harmonia", está se referindo à capacidade da IA em criar continuamente texto que seria naturalmente esperado de alguém que tivesse examinado minuciosamente o conteúdo presente em uma variedade de fontes, como páginas da web e outras fontes similares.

Imagine a digitalização de inúmeras páginas de texto criadas por humanos, seja na internet ou em livros digitalizados, com o objetivo de identificar todas as

instâncias de um determinado texto, e em seguida, determinar qual palavra segue esse texto e com que frequência. O ChatGPT realiza uma atividade comparável, embora ele não interprete o texto de maneira estrita; em vez disso, ele procura elementos que se relacionem de alguma forma em termos de significado. No entanto, o resultado final é a produção de uma lista organizada de palavras que podem vir a seguir, acompanhadas de suas respectivas "probabilidades" (WOLFRAM; 2023).

A ampla aceitação do ChatGPT por parte de usuários de diversas áreas demonstra o alcance e a versatilidade dessa ferramenta. As utilidades do ChatGPT vão desde auxiliar na elaboração de textos até agilizar a busca e recuperação de informações cruciais. Sua ascensão meteórica no cenário digital é testemunha de seu impacto significativo. Em questão de cinco dias desde o seu lançamento, a plataforma já contava com um milhão de usuários únicos. Em apenas dois meses, esse número cresceu exponencialmente, atingindo 100 milhões de usuários (CATALANO; LORENZI, 2023).

O sucesso do ChatGPT não passou despercebido pelos olhos atentos da mídia e do mercado financeiro. O fenômeno tornou-se um tópico de destaque em inúmeras reportagens, blogs e redes sociais, artigos de opinião e acadêmicos, gerando debates sobre o impacto da IA na sociedade (CATALANO; LORENZI, 2023).

No lado criativo, sistemas de IA texto-imagem como o DALL-E (também da OpenAI, lançado em 2021), *Stable Diffusion* (2022), *Midjourney* (2022) empurraram os limites do que a IA poderia alcançar em termos de geração de conteúdo visual. Esses sistemas poderiam pegar descrições textuais e transformá-las em imagens impressionantes, muitas vezes surreais e altamente imaginativas. Artistas e designers adotaram essas ferramentas como fontes de inspiração e como auxílio em seus processos criativos (MONDAL; DAS; VRANA, 2023).

O período de significativo investimento em IA generativa teve seu início com as gigantes da tecnologia, como Microsoft, Google e Baidu, assumindo a liderança nesse campo inovador. Essas empresas visionárias perceberam o potencial revolucionário das tecnologias de IA generativa e decidiram investir recursos

consideráveis para moldar o futuro digital. No entanto, não foram apenas as grandes corporações que embarcaram nessa jornada emocionante (MONDAL; DAS; VRANA, 2023).

Um fenômeno notável foi a entrada de diversas empresas menores na corrida para desenvolver modelos de IA generativa de ponta. Essas empresas, muitas vezes ágeis e especializadas em nichos específicos, reconheceram o poder transformador dessa tecnologia e decidiram competir no cenário global. À medida que o interesse e o investimento em IA generativa se expandiram, uma diversidade de inovações estão surgindo, impulsionando a evolução contínua desta área (REF, (MONDAL; DAS; VRANA, 2023).

A IA generativa tornou-se uma ferramenta poderosa em diversos campos, incluindo arte, escrita, desenvolvimento de software, jogos, marketing, dentre outros. Por exemplo, na arte, ela gerou obras visuais, música e literatura únicas, inspirando novas expressões criativas. Na escrita, auxiliou autores e criadores de conteúdo, sugerindo ideias e estilos específicos. No desenvolvimento de software, automatizou a geração e depuração de código, reduzindo erros. Na indústria de jogos, criou mundos virtuais realistas e personalizados. No marketing, gerou conteúdo rápido e direcionado, dentre outras diversas aplicações e campos (CATALANO; LORENZI, 2023) (MONDAL; DAS; VRANA, 2023).

No entanto, junto com essas promissoras aplicações, existem preocupações sobre o possível uso indevido da IA generativa. Essas preocupações incluem o uso da IA generativa para crimes cibernéticos, criação de *fake news* (notícias falsas) ou geração de *deepfakes*⁷ (WACH *et al.*, 2023). Essas tecnologias podem ser exploradas para enganar ou manipular indivíduos, apresentando desafios éticos e sociais que requerem consideração e regulamentação cuidadosas (CAO *et al.*, 2023).

À medida que a IA generativa continua a evoluir e penetrar em diversos setores, o equilíbrio entre seus benefícios e riscos potenciais permanece um aspecto crítico de seu desenvolvimento e implantação (CAO *et al.*, 2023). Lidar com essas

⁷ O termo "*deepfake*" é uma fusão das palavras "*deep learning*" (aprendizado profundo) e "*fake*" (falso), e representa uma técnica de síntese de imagens ou sons humanos que se baseia em avançadas técnicas de inteligência artificial. Essa tecnologia ganhou notoriedade por sua capacidade de combinar e manipular de maneira convincente elementos como áudio e vídeo, criando conteúdos aparentemente autênticos, mas que na verdade são fabricados (KIETZMANN *et al.*, 2020).

preocupações enquanto aproveita todo o potencial da IA generativa é um desafio que organizações, pesquisadores e formuladores de políticas estão ativamente trabalhando para enfrentar.

4.2.3 Perspectivas futuras

Segundo Chen *et al.* (2023), os avanços com as IA generativas, estão levando a IA em direção ao paradigma da IA Geral (AGI), o que significa que as máquinas poderão ser capazes de executar tarefas que anteriormente eram exclusivas do ser humano. Esses avanços têm implicações significativas para várias áreas, como saúde, educação, indústria, entretenimento e muito mais. Uma IA generalizada representa uma mudança fundamental na capacidade das máquinas de compreender, aprender e interagir com o mundo ao nosso redor. Para Fei *et al.* (2022), a capacidade de compreender e interpretar a linguagem humana é um aspecto essencial da inteligência humana, e os avanços nessa área permitem que as máquinas entendam e respondam a comandos, perguntas e interações em linguagem natural.

Apesar de não ser definido com precisão, é amplamente aceito que a AGI possui vários recursos principais, incluindo: (1) igualar ou exceder o desempenho humano em uma ampla classe de tarefas cognitivas (por exemplo, percepção, compreensão de leitura e raciocínio) em uma variedade de contextos e ambientes; (2) possuir a capacidade de lidar com problemas bem diferentes daqueles previstos por seus criadores; e (3) ser capaz de generalizar/transferir o conhecimento aprendido de um contexto para outro. Como podemos imaginar, conceber e obter um sistema AGI não apenas aceleraria a própria pesquisa de IA, mas também beneficiaria uma ampla gama de campos de IA+, incluindo neurociência, saúde e biomedicina (FEI *et al.*, 2022, p. 2, tradução nossa).

De acordo com Goertzel (2014), nos próximos vinte e cinco anos, é esperado que haja uma integração das abordagens de IA estreita e geral. Isso significa que os desenvolvedores estão criando agentes inteligentes capazes de interagir ativa e criativamente com ambientes complexos. No entanto, o autor argumenta que não há

uma justificativa convincente para concentrar todas as habilidades em um único indivíduo eletrônico ou em uma única IA.

Em vez disso, os recursos serão gradualmente utilizados para fortalecer as capacidades de IA projetadas para tarefas específicas. Essas IAs serão conectadas progressivamente para formar uma IA mais abrangente ou geral (GOERTZEL, 2014). Essa ideia é corroborada por um estudo da empresa de consultoria americana Gartner (2022), que observa que a criação de uma Inteligência Artificial Geral (AGI) é pouco provável nos próximos 10 anos, ou mesmo em um prazo mais longo. Portanto, a AGI provavelmente surgirá como resultado da combinação de diversas tecnologias de IA com finalidades específicas.

Apesar disso, a pesquisa nessa área continua em andamento, indicando que os avanços na IA estão ocorrendo, mas é mais provável que vejamos a evolução gradual e a combinação de tecnologias específicas para alcançar uma IA mais completa no futuro.

Essa transição tem o potencial de transformar profundamente a forma como interagimos com a tecnologia e como ela nos ajuda em nossa vida diária. Os algoritmos serão capazes de realizar tarefas que antes eram exclusivas do domínio humano, podendo nos levar a avanços científicos e tecnológicos significativos em diversas áreas. À medida que a AGI se torna uma realidade tangível, estamos diante da perspectiva de uma revolução que terá um impacto profundo em nossa relação com a tecnologia e como ela influenciará todos os aspectos de nossas vidas cotidianas (FEI *et al.*, 2022).

No entanto, é importante destacar que a transição para a AGI também levanta questões éticas e sociais cruciais, como privacidade, desigualdade, segurança cibernética e controle sobre as tecnologias avançadas. Portanto, é fundamental que a sociedade trabalhe em conjunto para desenvolver diretrizes e regulamentações que garantam que o uso da tecnologia seja usada para o bem da humanidade e que seus benefícios sejam distribuídos de forma justa e equitativa com foco na responsabilidade e liderança para maximizar seus benefícios e mitigar seus desafios.

5 INTELIGÊNCIA ARTIFICIAL E TEORIA ATOR-REDE EM DIÁLOGO

Neste capítulo, aprofundaremos nossa análise da IA sob a perspectiva crítica da Teoria Ator-Rede, explorando os agenciamentos tecnológicos e as controvérsias que envolvem esse campo. Sob essa perspectiva, a IA não é apenas uma ferramenta inerte, mas um ator com agência, suscetível às forças que a cercam. Portanto, adotaremos uma abordagem crítica para examinar as consequências sociais, éticas e políticas implicadas pela IA. Através da lente da TAR, nosso objetivo é trazer à luz algumas das complexidades e contradições inerentes a esse campo e promover uma compreensão mais profunda das implicações e oportunidades que a IA oferece à nossa sociedade.

De acordo com a Teoria Ator-Rede, o conceito de "agência" descreve a intrincada interação entre diversos elementos que desencadeiam ações e consequências no mundo. Essa ideia não se limita apenas ao domínio humano, estendendo-se a objetos, tecnologias e outros atores não humanos, todos participando ativamente em redes sociotécnicas (CAMILLIS; BUSSULAR; ANTONELLO, 2016).

Camillis, Bussular e Antonello (2016) destacam que a presença de uma agência é uma constante em cenários que envolvem transformações, ou seja, em situações em que indivíduos ou elementos exercem influência significativa sobre o seu entorno. Essa ideia ressalta a importância fundamental da agência, que se torna um fator determinante em contextos dinâmicos e em evolução constante.

Neste contexto, a agência pode ser entendida como a capacidade de atuar e produzir efeitos no ambiente circundante. Ela pode ser exercida por indivíduos, organizações, ou até mesmo por tecnologias e sistemas, e desempenha um papel central na condução das mudanças e na adaptação às transformações que ocorrem na sociedade e no mundo em geral (CAMILLIS; BUSSULAR; ANTONELLO, 2016).

A afirmação de Camillis, Bussular e Antonello (2016) também sugere que a agência não está restrita a um grupo específico de agentes, mas pode manifestar-se de diversas formas e em diferentes níveis, desde iniciativas individuais até movimentos sociais ou inovações tecnológicas de larga escala. A agência se torna uma força motriz que impulsiona as mudanças e, por sua natureza dinâmica, está

sempre presente em contextos onde ocorrem impactos e transformações. Portanto, compreender e reconhecer a agência é essencial para a análise e a gestão eficaz de processos de mudança em qualquer domínio.

Sob a perspectiva da TAR, a IA é concebida como uma complexa rede de atores, composta por cientistas, programadores, algoritmos, dados e modelos cognitivos. O desenvolvimento da IA vai além da simples programação de máquinas e inclui também a sua interação com o mundo real e a colaboração entre seres humanos e algoritmos. Essa abordagem holística destaca a importância das relações e conexões entre os diversos agentes envolvidos na IA, abrangendo não apenas os aspectos técnicos, mas também as implicações éticas, sociais e políticas.

Latour (2012) promove a ideia de "simetria" ao tratar tanto seres humanos quanto não humanos de forma igualitária na análise das redes sociotécnicas, as quais se moldam por meio de conflitos e negociações. É crucial reconhecer que fatores externos influenciam o desenvolvimento da IA e adotar uma postura crítica em relação às narrativas simplistas de progresso tecnológico. Conceber a IA como parte integrante de uma ampla rede sociotécnica nos instiga a refletir sobre as implicações éticas, sociais e políticas inerentes ao seu desenvolvimento e uso.

Nesse contexto, o agenciamento se concretiza por meio das interações e associações entre diversos atores, criando uma rede na qual a IA é concebida, implementada e utilizada. Isso vai além dos aspectos meramente técnicos, incorporando considerações de ordem ética e social. Essa abordagem nos permite compreender a complexidade intrincada da IA, reconhecendo que ela resulta da interação de uma variedade de atores e está em constante evolução. Isso promove uma mentalidade de aprendizado contínuo e adaptativo. Portanto, a IA se configura como um fenômeno híbrido, combinando ação humana e não humana em novas possibilidades mediadas pela colaboração de uma ampla rede de agentes heterogêneos.

A história da IA é uma narrativa fascinante e intrincada que se desenrolou ao longo de décadas, sendo moldada por uma complexa teia de atores e eventos cruciais. Ao analisarmos essa jornada sob a ótica da TAR, somos levados a uma compreensão mais profunda dos agenciamentos que deram forma à IA como a

conhecemos hoje, foi moldada por uma série de atores que desempenharam papéis cruciais na rede sociotécnica em torno da IA.

O campo da IA tem Alan Turing como seu patrono. Seu trabalho seminal, "Máquinas de Computação e Inteligência", publicado em 1950, introduziu conceitos fundamentais da IA ao questionar se as máquinas podem pensar. O Teste de Turing, conhecido como o "Jogo da Imitação", propôs uma maneira de avaliar a capacidade de uma máquina em exibir comportamento inteligente indistinguível do comportamento humano (RUSSEL; NORVIG, 2013). Essa ideia influenciou significativamente o desenvolvimento da IA, mostrando como os atores humanos podem moldar a direção da pesquisa.

Em 1956, a Conferência de Verão em Dartmouth College, liderada por John McCarthy, influenciado pelas ideias de Turing, é amplamente considerada como o marco zero do agenciamento da IA. Durante o evento, o termo "Inteligência Artificial" foi cunhado, e uma proposta foi apresentada: a realização de um estudo de dois meses com uma equipe de dez homens sobre o tópico da IA (MCCORDUCK, 2004). Evidenciando como os atores humanos, como McCarthy e sua equipe, desempenharam um papel vital na consolidação da IA como campo de estudo acadêmico e como as ideias de Turing se propagaram na rede sociotécnica.

Simon, por exemplo, percebeu que as máquinas poderiam ter um propósito mais amplo do que apenas processar números. Sua observação de como os computadores imprimiam palavras em um mapa para sistemas de defesa aérea sugeriu que essas máquinas poderiam ser usadas para lidar com imagens, caracteres e símbolos, pavimentou o surgimento da IA simbólica e expandindo significativamente o escopo da IA (TAULLI, 2019). Através da lente da TAR, Simon emerge como um ator central que contribuiu para a ascensão da IA simbólica, estabelecendo a mesma como domínio no campo de desenvolvimento da IA ao longo de um período significativo.

Marvin Minsky também desempenhou um papel importante na história da IA ao desenvolver a *Stochastic Neural Analog Reinforcement Calculator* (SNARC), a primeira máquina de rede neural artificial em 1951. Essa máquina, que operava com o sistema operacional DOS, representou um avanço significativo ao explorar a

capacidade de aprendizado de máquinas (BARBOSA; BEZERRA, 2020). As experiências de Minsky com a SNARC demonstram como os atores impulsionam a pesquisa e a experimentação de diferentes formas de desenvolvimento de IA junto a objetos técnicos.

O Logic Theorist, apresentado durante a conferência, foi o primeiro programa de computador de IA e representa um desses atores. Desenvolvido na Corporação de Pesquisa e Desenvolvimento (RAND) por Allen Newell, Cliff Shaw e Herbert Simon, o programa focava na resolução de problemas matemáticos a partir do livro "Principia Mathematica", escrito por Alfred Whitehead e Bertrand Russell. No entanto, para tornar isso possível, eles não apenas criaram o Logic Theorist, mas também desenvolveram uma linguagem de alto nível chamada IPL (Information Processing Language), que se tornou fundamental para muitos projetos de IA posteriores (TAULLI, 2019). Ilustrando como os atores não humanos, como programas de computador e linguagens de programação, têm um papel ativo na formação da IA.

Além disso, a memória limitada do computador IBM 701 com o qual trabalhavam apresentou um desafio significativo. Para superá-lo, Newell, Shaw e Simon desenvolveram o conceito de processamento de listas, que permitiu a alocação e desalocação dinâmica de memória durante a execução do programa (TAULLI, 2019). Demonstrando como a evolução das tecnologias é uma parte intrínseca da história da IA, com atores não humanos se adaptando para superar limitações técnicas.

A TAR nos permite compreender o marco-zero da IA como o agenciamento de uma nova rede complexa de relações entre atores humanos e não humanos. Os atores não humanos, como programas de computador e linguagens de programação, desempenharam papéis ativos na evolução tecnológica, enquanto os atores humanos, como Turing, McCarthy, Simon, Minsky e outros, moldaram a direção da pesquisa e a aplicação da IA. Essa perspectiva nos ajuda a apreciar a interconexão dinâmica e evolutiva que caracteriza esse momento da história.

Claramente, o período que abrangeu os anos de 1956 a 1974 ficou marcado na história da IA como a "Era de Ouro da IA" (TAULLI, 2019). Nesse cenário

efervescente, a TAR oferece uma lente analítica perspicaz para entender como, novamente, diversos atores interagiram e moldaram o desenvolvimento da IA.

Em primeiro lugar, as tecnologias computacionais emergiram como atores nesse cenário. Os sistemas baseados em tubos de vácuo cederam espaço a sistemas menores e mais ágeis, equipados com circuitos integrados avançados e maior capacidade de armazenamento (TAULLI, 2019). A evolução das tecnologias computacionais, com suas crescentes capacidades de processamento, pode ser vista como um ator que habilitou o desenvolvimento de algoritmos e sistemas mais avançados de IA.

Outro ator relevante neste período foi a Agência de Projetos de Pesquisa Avançada (ARPA), que desempenhou um papel crucial ao fornecer financiamento para pesquisas em IA. A ARPA pode ser vista como um ator que contribuiu para a rede de pesquisa em IA, fornecendo recursos financeiros que viabilizaram muitos projetos. Seus projetos financiados tinham poucas restrições, buscando fomentar a inovação revolucionária. Universidades de renome, como Stanford, MIT e Carnegie Mellon, foram beneficiadas por esse financiamento (TAULLI, 2019), tornando-se também atores na rede de pesquisa em IA.

Já os pesquisadores acadêmicos também são atores na formação da IA durante essa época. Grande parte da inovação em IA foi conduzida por esses pesquisadores, em universidades e empresas beneficiadas pelo financiamento da ARPA (TAULLI, 2019), que representa um agenciamento de atores ativos que contribuíram com conhecimento e desenvolveram novas abordagens na IA.

No entanto, a "Era de Ouro" também foi marcada por um conflito de ideias notável, representado por Minsky e Rosenblatt. Minsky defendia a abordagem simbólica (lógica computacional), enquanto Rosenblatt promovia o conexionismo (o que viria a se tornar mais tarde o aprendizado profundo) (TAULLI, 2019). Esses dois atores influenciaram diretamente a direção da pesquisa em IA, com suas críticas e ideias moldando o campo de maneira significativa se revelou como um fator significativo que influenciou a direção da pesquisa em IA, pois houve um aumento significativo em pesquisas relacionadas à IA simbólica na rede, resultando em

desestímulo por parte dos acadêmicos no desenvolvimento da IA conexionista (TAULLI, 2019).

Além disso, as transformações na IA ao longo desse período são fruto das interações dinâmicas entre todos esses atores. Novas abordagens e metodologias continuaram a ser desenvolvidas à medida que pesquisadores, tecnologias e ideias interagem. A TAR nos ajuda a compreender como essas interações constantes influenciaram a evolução da IA.

No contexto da IA, fatores tecnológicos, econômicos, políticos e sociais desempenham papéis significativos em seu desenvolvimento. Isso pode ser explicado por meio de uma tendência recorrente na comunidade de IA que é a alternância entre períodos de grande otimismo, conhecidos como a "Primavera da IA", e períodos de escassez de recursos financeiros e pessimismo, chamados de o "Inverno da IA" (MITCHELL, 2019).

Na perspectiva da TAR, esses ciclos podem ser compreendidos como resultado das interações entre diversos atores envolvidos no campo da IA. Esses atores incluem pesquisadores, financiadores, a mídia e a própria tecnologia da IA. O otimismo é alimentado pela promessa de avanços tecnológicos revolucionários, enquanto a escassez muitas vezes ocorre devido à falta de financiamento, expectativas irrealistas e ceticismo, também podem ser considerados atores na rede, uma vez que influenciaram a trajetória da IA ao longo do tempo.

Durante os períodos de "invernos da IA" podemos ter uma compreensão mais profunda da dinâmica da IA ao longo do tempo, destacando como essa tecnologia é influenciada por uma ampla gama de atores que interagem e moldam seu desenvolvimento por meio de suas ações na rede. O fenômeno do exagero em torno da IA também pode ser entendido como atores afetam a percepção e a adoção da tecnologia. Desempenhando papéis de amplificação de expectativas exageradas, muitas vezes levando a uma queda na reputação da IA (MITCHELL, 2019).

No entanto, apesar das quedas na reputação, a IA continuou a se desenvolver com sucesso e a produzir novas inovações (NEWQUIST, 2020). Isso pode ser interpretado como uma demonstração da resiliência da IA como um ator

em rede. Mesmo quando a percepção pública é negativa, a IA continua a se conectar e a influenciar outros atores na rede.

Este contexto nos leva a uma compreensão mais aprofundada da complexa dinâmica de poder que envolve a IA. Ela é moldada e influenciada por uma ampla gama de atores que interagem de maneira interconectada, influenciando sua trajetória de desenvolvimento e integração na sociedade. Essa abordagem nos ajuda a apreciar a IA não apenas como uma tecnologia isolada, mas como um ator que desafia e transforma diversas esferas de nossas vidas.

A TAR nos mostra que diferentes atores podem estar em disputa, cada um promovendo seus próprios interesses (LATOURE, 2012). Também enfatiza a importância de reconhecer diferentes formas de tecnologia (AKRICH, 2014). O relatório de Lighthill que descredibilizou a IA não reconheceu o potencial da IA fraca (TAULLI, 2019). Isso ilustra como a falta de reconhecimento de diferentes abordagens da IA pode afetar negativamente seu desenvolvimento. O debate público televisionado entre Lighthill e outros pesquisadores, no qual ele apresentou argumentos pessimistas sobre a IA (TAULLI, 2019), também destaca como diferentes visões podem moldar a percepção pública da tecnologia.

O "primeiro inverno da IA" teve um impacto direto na capacidade de desenvolvimento da IA na época, levando muitos pesquisadores a mudar de carreira. A adaptação das redes de atores ao longo do tempo, em resposta ao descrédito e às desconfianças trazidas pelo inverno da IA, é evidenciada pela transição para termos como "aprendizado de máquina" e "redes neurais" (TAULLI, 2019). Isso destaca como as redes de atores podem se adaptar a desafios significativos ao empregar novas terminologias para continuar sua pesquisa, demonstrando a resiliência da rede diante de adversidades.

Esse cenário nos ajuda a enxergar a IA não como uma entidade isolada, mas como parte de uma rede complexa de atores humanos e não humanos, influenciada por fatores sociais, econômicos e políticos. Ao adotar essa perspectiva, podemos compreender melhor por que a IA passou por períodos de avanço e estagnação ao longo do período, e como essa tecnologia continua a evoluir em resposta a mudanças na rede de atores que a moldam. Podemos observar também, como

diferentes atores, como o governo, os pesquisadores acadêmicos, a mídia e a própria opinião pública, possuem agendas diferentes que influenciam o desenvolvimento da IA.

Durante a rápida e curta ascensão dos sistemas especialistas e o "segundo inverno da IA", podemos compreender como a rede complexa de atores humanos (especialistas em suas áreas de atuação) e não humanos (computadores e programas de software) colaboram e coevoluem. Os sistemas especialistas, como o DENDRAL e o MYCIN, foram pioneiros em IA. Esses sistemas baseiam-se em um vasto conjunto de regras específicas obtidas de especialistas de domínio (RUSSEL; NORVIG, 2013). Sob a perspectiva da TAR, essas regras representam o conhecimento humano sendo hibridizado na rede.

Ao longo dos anos, os sistemas especialistas evoluíram em resposta ao avanço da capacidade de processamento dos computadores (TAULLI, 2019). A coevolução entre os atores na rede também desempenhou um papel importante na adaptação contínua desses sistemas às mudanças nas necessidades e demandas do mercado. À medida que novos desafios e oportunidades surgiram, os especialistas e os desenvolvedores de software colaboraram para aprimorar os sistemas existentes e criar novas soluções.

Entretanto, os sistemas especialistas não estavam imunes a desafios. As complicações associadas à incorporação de dados precisos e à gestão de sua crescente complexidade evidenciam a natureza dinâmica da tecnologia, sujeita a controvérsias e contingências (TAULLI, 2019). A manutenção e atualização de sistemas especialistas representam processos intrincados, nos quais diversas partes interconectadas em uma rede dinâmica de atores frequentemente influenciam e, por vezes, complicam o desenvolvimento dessas soluções.

O declínio temporário na popularidade dos sistemas especialistas, no final da década de 1980 (TAULLI, 2019), reverbera como fatores sociais e econômicos influenciam a aceitação e o sucesso da tecnologia. Resultando em fusões e falências de empresas (NEWQUIST, 2020). Ressaltando como a aceitação e o sucesso dessas tecnologias estão sujeitas a fatores sociais e econômicos, não

apenas a características intrínsecas da tecnologia. A conjuntura econômica e as preferências do mercado também são atores relevantes na rede sociotécnica.

No entanto, é importante notar que os sistemas especialistas continuaram a exercer influência duradoura. Mesmo após seu declínio, esses sistemas moldaram outras áreas de pesquisa em IA, como a aprendizagem de máquina e a IA distribuída (RUSSEL; NORVIG, 2013). Isso demonstra como a rede de atores, incluindo os sistemas especialistas, desempenha um papel duradouro na construção da rede da IA.

Sob a ótica da TAR, podemos compreender que a evolução das tecnologias de IA estão sujeitas a contingências sociais, tecnológicas e econômicas. Essa abordagem nos ajuda a compreender melhor a dinâmica por trás do desenvolvimento e do declínio temporário dessas tecnologias, bem como sua influência contínua na pesquisa e desenvolvimento em IA. A análise crítica desses períodos nos oferece uma visão de rede do desenvolvimento da IA. Isso nos ajuda a entender melhor a dinâmica ao longo dos períodos de "invernos" e "primaveras", bem como sua contínua influência na pesquisa e na sociedade como um todo.

No contexto do agenciamento que promoveu o ressurgimento da IA após seu período de estagnação, Hinton emergiu como um ator principal. Suas ações e ideias desempenharam um papel fundamental na moldagem da evolução das redes neurais, e, por conseguinte, no rumo da IA. Essencialmente, Hinton defendeu a revitalização das redes neurais, originalmente desenvolvidas por Rosenblatt, com uma visão ousada: a de que essas redes poderiam aprender de maneira semelhante aos seres humanos, através do reconhecimento de padrões. Esses elementos, conforme destacados por Newquist (2020), desempenharam um papel crucial no renascimento da IA.

Hinton desafiou as visões predominantes da época, defendendo a capacidade das redes neurais de aprender através de treinamento inicial, tentativa e erro e uso de processadores não convencionais, GPUs, em vez de unidades centrais de processamento (CPUs) (NEWQUIST, 2020). Sob a perspectiva da TAR, esse movimento desempenhou um papel fundamental na transformação da IA, servindo como um agenciamento de híbridos na rede.

Ademais, a introdução das GPUs (unidades de processamento gráfico) no campo da IA é outro marco importante. As GPUs são atores não humanos que aceleraram significativamente o treinamento de redes neurais, tornando-o mais rápido e eficiente (NEWQUIST, 2020). Exemplificando como a tecnologia é uma parte intrínseca da rede de atores na IA.

No entanto, é essencial notar que a resistência e a desconfiança manifestadas por figuras proeminentes como Minsky e Papert representaram atores que se opuseram à visão de Rosenblatt durante as fases iniciais do desenvolvimento da IA, chegando até a desacreditar as redes neurais em seu influente livro "*Perceptrons*" (NEWQUIST, 2020). Essa divergência de opiniões ilustra vividamente o caráter intrinsecamente controverso e disputado que permeou o desenvolvimento da IA. Quando consideramos a perspectiva da TAR, Minsky, Papert e até mesmo o mencionado livro são atores cujas ações e argumentos desempenharam papéis significativos na configuração da complexa rede de controvérsias que envolvem o campo da IA.

A capacidade de processamento dos computadores desempenhou um papel crucial no progresso da IA, em conjunto com a viabilidade dos dados (NEWQUIST, 2020). Esses dois fatores foram fundamentais para viabilizar o avanço da IA. Neste cenário, chips de computador e dados, também desempenham um papel significativo na construção da rede de atores da IA. Isso exemplifica como atores não humanos estão intrinsecamente envolvidos na construção da IA.

A competição ImageNet, outro agenciamento, iniciada pela pesquisadora Fei-Fei Li, é um exemplo de como a comunidade científica e acadêmica desempenha um papel fundamental na construção da IA. Li liderou uma equipe que criou um banco de dados de imagens vasto e diversificado, envolvendo a participação de muitos atores humanos que marcaram e prepararam os dados (NEWQUIST, 2020). A competição serviu como um espaço onde diferentes atores puderam testar suas abordagens e algoritmos de IA.

A transição da primeira onda da IA para a segunda onda moderna ressalta a mudança de ênfase da IA simbólica para a IA conexionista (GOODFELLOW; BENGIO; COURVILLE, 2016). Isso ilustra claramente como as tendências e

paradigmas na IA são moldados por atores que promovem distintas abordagens e estratégias.

As *big techs* (grandes empresas de tecnologia), como Google, Amazon, Apple, IBM e Baidu, também desempenham papéis importantes na evolução da IA, através de investimentos em pesquisa, aquisições de *startups* de IA e coleta de dados (NEWQUIST, 2020). Essas empresas são atores influentes na rede da IA.

Ao analisar a evolução da IA neste período, destaca-se a formação de uma nova rede complexa de atores que interage, influencia e constrói essa tecnologia em constante desenvolvimento. Cada ator, seja um pesquisador, algoritmo, *hardware* ou conjunto de dados, desempenha um papel ativo na construção da IA, e a compreensão dessa dinâmica é essencial para apreciar o cenário atual da IA.

Sob a perspectiva da TAR, é possível analisar a ascensão do ChatGPT e o subsequente *hype* (furor) em torno de outras IAs generativas como um agenciamento complexo de híbridos. Nesse contexto, vários elementos se entrelaçam e interagem para moldar a realidade tecnológica e, por conseguinte, os impactos que essa tecnologia tem sobre a sociedade.

O ChatGPT, desenvolvido pela OpenAI, é um exemplo notável de como a tecnologia pode agir como um ator híbrido na rede. Ele não é apenas uma ferramenta de IA, mas uma construção que reflete os elementos de treinamento, dados da internet e desenvolvimento humano. Os LLM, como o GPT-3 e o GPT-4, são atores não-humanos que desempenham o papel de hibridização com a geração de respostas autônomas e na ampliação das capacidades de processamento de linguagem natural por meio da extração e interpretação de dados gerados por humanos.

Além do mais, o *hype* em torno do ChatGPT e outras IA generativas em 2022 agiu como um mecanismo de amplificação para o agenciamento da IA na rede. Empresas, pesquisadores, investidores e governos foram atraídos pela visibilidade e pelo potencial da IA. Isso resultou em investimentos crescentes e um aumento na pesquisa e desenvolvimento em IA, impulsionando mais avanços científicos e tecnológicos (DWIVEDI et al., 2023).

A integração da IA em produtos e serviços pelas *big techs* (NEWQUIST, 2020) demonstram como as tecnologias de IA são incorporadas em redes mais amplas. Essa integração não apenas impulsiona a inovação, mas também reforça o agenciamento e a interconexão entre atores humanos e não humanos na rede.

Sob a perspectiva da TAR, podemos entender como esses diversos elementos estão interconectados e como suas interações complexas influenciam o desenvolvimento, a adoção e o impacto da tecnologia. Ela também nos permite analisar as controvérsias, os desafios éticos e as questões sociais que emergem à medida que esses agenciamentos evoluem, ajudando-nos a obter uma compreensão mais profunda da interseção entre tecnologia e sociedade. Portanto, a ascensão da IA é um fenômeno que transcende simplesmente a inovação tecnológica e se inserem em uma rede complexa de influências e atores em constante interação.

No entanto, enquanto admiramos os avanços tecnológicos e os benefícios que a IA trouxe, também não podemos ignorar as críticas que surgiram em relação a esse processo. Um dos principais pontos de crítica é o poder considerável que as *big techs* acumularam (GALLOWAY, 2017). Com sua capacidade de moldar o uso da IA em diversos setores, essas empresas tornaram-se gigantes, muitas vezes dominando mercados inteiros e sufocando a competição (ZUBOFF, 2019). Isso levanta preocupações legítimas sobre a concentração de poder e a falta de diversidade e inovação em um cenário dominado por poucos (CHANG, 2018).

Além disso, a coleta maciça de dados por parte das *big techs* para alimentar algoritmos de IA levanta questões de privacidade e segurança (ZUBOFF, 2019). O risco de violações de dados e o potencial para a exploração indevida desses dados são preocupações constantes (SCHNEIER, 2016). Um dos principais desafios relacionados à IA é a preservação da privacidade. À medida que sistemas de IA coletam, analisam e armazenam enormes quantidades de dados pessoais, cresce a preocupação sobre como essas informações são utilizadas e protegidas (HANSON, 2016). A privacidade individual torna-se um direito fundamental que deve ser protegido, ao mesmo tempo em que se aproveita o potencial benéfico da IA (FRISCHMANN; SELINGER, 2018).

A IA frequentemente baseia seu aprendizado em dados históricos, o que significa que ela pode incorporar os preconceitos presentes nesses dados (O'NEIL, 2020). Isso suscita preocupações éticas significativas, já que a IA pode atuar como um agente que amplifica desigualdades preexistentes (NOBLE, 2018). Um exemplo notório desse problema é o viés algorítmico, como evidenciado no caso dos algoritmos de reconhecimento facial que demonstraram viés racial (BUOLAMWINI; GEBRU, 2018) e nos sistemas automáticos de seleção de currículos que podem perpetuar desigualdades de gênero (AUMÜLLER-WAGNER, 2019). Além disso, isso ressalta como o hibridismo nas redes afeta até mesmo os não humanos, ao perpetuar valores humanos nos algoritmos. Portanto, torna-se essencial desenvolver estratégias eficazes para mitigar e corrigir esses vieses, com o objetivo de promover a equidade e a justiça. Uma abordagem crucial é o desenvolvimento de algoritmos mais justos, transparentes e responsáveis.

A desigualdade digital também é uma preocupação crescente, já que nem todos têm acesso igualitário às tecnologias de IA, aprofundando a divisão entre aqueles que podem se beneficiar dessas inovações e aqueles que não podem (ZUBOFF, 2019). Além disso, a automação impulsionada pela IA pode resultar na perda de empregos (MCKINSEY; 2017, 2021) (OCDE, 2023), levantando questões sobre o impacto social e econômico dessa transição (BRYNJOLFSSON; MCAFEE, 2014).

A ética e a responsabilidade no uso da IA também estão no centro das preocupações (COECKELBERGH, 2020). A falta de transparência nas decisões algorítmicas, também conhecida como opacidade algorítmica, cria um ambiente propício para a formação de "caixas pretas" em torno da IA como técnica, o que significa que os processos de tomada de decisão algorítmica são obscurecidos e difíceis de serem compreendidos pelos indivíduos que interagem com eles e a ausência de regulamentação eficaz dificultam a responsabilização das empresas por danos causados por sistemas algoritmos (PASQUALE, 2015).

Além disso, essa crescente atenção à IA trouxe à tona questões éticas e regulatórias, destacando a importância da responsabilidade compartilhada no agenciamento da rede (COECKELBERGH, 2020). A sociedade como um todo se

tornou um ator importante nesse contexto, exigindo a consideração de questões relacionadas ao viés, privacidade e segurança.

Também exercem um papel importante na rede, muitas priorizando seus próprios interesses sobre o bem-estar da sociedade. A influência das *big techs* nas políticas e regulamentações governamentais levantam questões sobre um possível conflito de interesses e a influência desproporcional na formulação de políticas (GALLOWAY, 2017). No entanto, a TAR também pode servir de alerta para esses desafios. À medida que a IA se torna mais difundida, é fundamental que a rede de atores colabore para desenvolver regulamentações adequadas e diretrizes éticas que garantam um uso responsável e justo da tecnologia.

Entretanto, juntamente com as promissoras aplicações da IA generativa, surgem preocupações complexas sobre seu potencial uso indevido, engendrando uma rede intrincada de atores e elementos que desempenham papéis interconectados na definição da percepção pública e da regulamentação dessa tecnologia. Essas inquietações abrangem uma série de áreas, como o uso da IA generativa em cibercrimes, na disseminação de notícias falsas (*fake news*) e na criação de *deepfakes* (WACH et al., 2023).

A reflexão do desenvolvimento da IA sob perspectiva da TAR nos conduz a uma profunda reavaliação da IA, que vai além da visão simplista de um mero avanço tecnológico isolado. Em vez disso, nos impulsiona a enxergar a IA como um resultado complexo e, por vezes, caótico, emergindo de interações intrincadas entre diversos atores, sejam eles humanos ou não humanos.

Desde sua concepção inicial por Turing, o agenciamento em Dartmouth com a apresentação do *Logic Theorist*, seus respectivos invernos e primaveras, o retorno da IA conexionista até a atual era do ChatGPT, a IA tem sido moldada por uma teia de atores complexa. No entanto, a TAR nos obriga a questionar: essa complexidade está realmente resultando em avanços significativos? Ou está apenas criando um emaranhado confuso de influências e agendas que frequentemente se sobrepõem e entram em conflito?

Além disso, a rápida disseminação da IA levanta questões éticas, de privacidade e de segurança que desafiam a capacidade da comunidade de IA de

lidar de forma responsável (COECKELBERGH, 2020) com seu próprio agenciamento. É inegável que a IA teve avanços notáveis desde o seu marco zero, mas a TAR nos convida a olhar para além dos avanços tecnológicos. Devemos questionar se a complexa rede de atores está, de fato, promovendo uma IA verdadeiramente benéfica para a humanidade ou se está apenas gerando um ciclo incessante de *hype* e desilusão.

À medida que a IA se torna mais integrada em nossa sociedade, é crucial que adotemos uma abordagem crítica, inspirada pela TAR, para entender as complexas redes de atores que a impulsionam. Devemos questionar não apenas o que a IA pode fazer, mas também quem se beneficia dela, quem a controla e quem sofre suas consequências. Somente assim poderemos abordar plenamente as implicações éticas, sociais e políticas dessa tecnologia em constante evolução.

É imperativo que mantenhamos essa perspectiva crítica à medida que a IA se torna uma parte cada vez mais ubíqua e intrínseca de nossas vidas e sociedade. A TAR nos lembra que, por trás das inovações brilhantes e das promessas de transformação, há complexas dinâmicas de poder e interesses em jogo.

Portanto, a IA não é apenas uma ferramenta neutra; ela reflete e amplifica os valores e intenções de seus criadores e controladores. É fundamental questionar quem define os parâmetros da IA, quais dados são usados para treiná-la e quem decide como ela é aplicada. Devemos garantir que essas decisões não sejam monopolizadas por um grupo seleto, mas sim moldadas por uma ampla gama de vozes e perspectivas, a fim de evitar a concentração de poder e potenciais consequências negativas.

A TAR nos desafia a não apenas aceitar a IA como um inevitável avanço tecnológico, mas a considerá-la como uma questão central da sociedade contemporânea que requer um exame crítico constante. Devemos estar preparados para ajustar nossa abordagem à medida que novos desafios e dilemas éticos surgirem.

5.1 MEDIANDO A CONTROVÉRSIA DO FUTURO DO TRABALHO

Neste tópico, vamos aprofundar nossa análise de controvérsias envolvendo a IA sob a perspectiva da TAR envolvendo a questão do trabalho. Reconhecemos a intrincada teia de atores que influencia e é influenciada pela rede da IA, e compreendemos as implicações profundas que essa abordagem traz para a nossa compreensão da IA e suas interações com a sociedade e a cultura.

A TAR nos instiga a considerar a IA como um agente dinâmico em constante evolução, tanto moldando quanto sendo moldada por sua rede de interações. Isso nos leva a refletir sobre como a IA pode continuar a se desenvolver e como as controvérsias que a cercam podem se transformar ao longo do tempo.

Na obra "Ciência em Ação", Latour (2000) apresenta uma perspectiva transformadora sobre a construção do conhecimento científico e tecnológico, enfatizando o papel central das controvérsias nesse processo. Ele desafia a visão tradicional de que a ciência e a tecnologia são produtos de um processo objetivo de descoberta e argumenta, em vez disso, que são produtos de complexas negociações sociais e debates envolvendo diversos atores.

Na análise de Latour (2000), as controvérsias são cenários em que uma variedade de grupos sociais, incluindo cientistas, engenheiros, políticos, grupos de interesse e outros intervenientes, disputam questões relacionadas à ciência e tecnologia. Essas controvérsias podem abranger desde desacordos sobre fatos científicos até dilemas éticos, interesses econômicos e decisões políticas. Para Latour, é exatamente por meio dessas controvérsias que o conhecimento científico e tecnológico toma forma e, eventualmente, se estabiliza.

Latour (2000) ressalta que as controvérsias desempenham um papel crucial na moldagem do conhecimento científico e tecnológico, pois nesses momentos, diferentes atores sociais estão ativamente envolvidos em negociações e debates que, em última análise, definem o que é aceito como verdade dentro da comunidade científica e da sociedade em geral. Isso enfatiza a importância de compreender a ciência e a tecnologia como atividades profundamente enraizadas em contextos sociais e culturais, e não como processos objetivos e desvinculados da influência humana.

A IA está desempenhando um papel fundamental na aceleração da transformação das atividades laborais em inúmeros setores e indústrias. Esse avanço tecnológico está gerando um impacto profundo em nossa sociedade, suscitando uma série de questões cruciais que merecem nossa atenção.

As implicações sociais da IA também incluem preocupações sobre emprego e desigualdade (BRYNJOLFSSON; MCAFEE, 2014). A automação impulsionada pela IA pode resultar na substituição de empregos tradicionais, exigindo uma adaptação da força de trabalho e políticas de inclusão social para evitar a marginalização de certos grupos (CHUI ; MANYIKA; MIREMADI, 2016).

A rápida adoção da IA levanta preocupações significativas sobre o futuro do emprego (BESSEN, 2018). À medida que algoritmos avançados e sistemas automatizados ganham terreno em tarefas tradicionalmente realizadas por seres humanos, surge a preocupação com a automação de empregos e a potencial substituição de trabalhadores por máquinas (OCDE, 2023).

Novos modelos de negócios baseados em dados e automação estão surgindo, alterando a dinâmica competitiva em várias indústrias (BRYNJOLFSSON; MCAFEE, 2014). A transformação digital está gerando eficiência, mas também desafios em termos de regulação, segurança de dados e equidade econômica (WEST, 2018). Como sociedade, devemos enfrentar esses desafios para garantir que a economia continue a beneficiar a todos.

A capacidade da IA de permitir que sistemas executem uma ampla gama de tarefas, que anteriormente eram consideradas exclusivas do domínio humano, é notável (RUSSEL; NORVIG, 2013). Essas tarefas variam desde atividades repetitivas, regidas por regras preestabelecidas, até tarefas mais sofisticadas, que envolvem a análise de dados, a tomada de decisões complexas e até mesmo a manifestação de criatividade (GOODFELLOW; BENGIO; COURVILLE, 2016). Esse cenário instiga reflexões profundas acerca do número de empregos que podem ser substituídos pela automação e como isso impactará a taxa de desemprego (BRYNJOLFSSON; MCAFEE, 2014).

No contexto da automação, torna-se evidente que muitos postos de trabalho podem, de fato, ser suplantados por máquinas e algoritmos. Isso levanta

preocupações significativas sobre a sustentabilidade das ocupações tradicionais e a necessidade de adaptação dos trabalhadores para novas realidades econômicas (FREY; OSBORNE, 2017).

Em 2017, o McKinsey Global Institute, uma empresa de consultoria empresarial sediada nos Estados Unidos, lançou um estudo intitulado "Jobs lost, jobs gained: workforce transitions in a time of automation" (Empregos perdidos, empregos ganhos: transições na força de trabalho em tempos de automação). Esse estudo representou um exame detalhado dos possíveis desdobramentos do mundo do trabalho em um período que abrange desde 2015 até 2030. Suas conclusões apontaram para a inevitabilidade de mudanças dramáticas nos processos produtivos e, por extensão, na sociedade como um todo.

A pesquisa destacou o impacto profundo que a automação, juntamente com outras tecnologias como a IA, terá nas economias globais e na forma como as pessoas desempenham suas funções no mercado de trabalho. Uma das conclusões mais notáveis foi a estimativa de que mais de 50% de todos os empregos no mundo poderiam ser automatizados até o ano de 2030. Essa projeção trouxe consigo uma série de implicações significativas que se estendem por diversos setores e regiões do planeta (MCKINSEY, 2017).

Essa mudança radical no cenário do emprego levanta questões prementes sobre como as sociedades e as economias devem se adaptar a essa transição. O estudo sublinhou a importância de planejar estratégias de reskilling (requalificação profissional) para trabalhadores cujas ocupações são particularmente suscetíveis à automação, a fim de minimizar o impacto negativo sobre o desemprego e o subemprego. Além disso, enfatizou a necessidade de políticas públicas inovadoras, parcerias entre setor público e privado, bem como o desenvolvimento de novos modelos de negócios que possam aproveitar os benefícios da automação sem deixar ninguém para trás (MCKINSEY, 2017).

O estudo lançou luz sobre o futuro do trabalho em uma era de automação e tecnologia avançada, destacando a urgência de adaptação e preparação para as transformações iminentes no mercado de trabalho e na sociedade em geral. Essas transformações no mundo do trabalho, impulsionadas pelo advento de tecnologias

disruptivas e acelerada pela pandemia da Covid-19, é um fenômeno que merece uma análise mais aprofundada. Por isso, um estudo complementar foi lançado em 2021, intitulado "O Futuro do Trabalho Pós-COVID-19," forneceu percepções valiosas sobre como as novas tecnologias e as circunstâncias excepcionais da pandemia estão moldando o cenário laboral global (MCKINSEY, 2021).

Primeiramente, é crucial reconhecer que o advento de novas tecnologias de informação estão redefinindo o trabalho de forma fundamental. Tecnologias de IA, aliadas à automação avançada e a conectividade em rede estão transformando não apenas a natureza das ocupações, mas também a maneira como elas são executadas. A digitalização do trabalho tornou-se uma necessidade, e os trabalhadores agora dependem ainda mais de sistemas digitais para desempenhar suas funções, muitas vezes de maneira remota. A automatização de tarefas de baixa complexidade é outra consequência dessa mudança, levando a uma reconfiguração das funções tradicionais (MCKINSEY; 2017, 2021).

A pandemia da Covid-19 foi um catalisador importante nesse processo. As restrições de confinamento e o distanciamento social forçaram empresas e indivíduos a adotarem soluções tecnológicas para manter a continuidade dos negócios e o trabalho remoto. Isso, por sua vez, acelerou a aceitação da automação e da IA em muitos setores, resultando em uma rápida evolução das tendências previstas em estudos anteriores. A necessidade de minimizar a proximidade física no trabalho impulsionou a automação de processos e a implementação de soluções de IA em várias indústrias, desde a manufatura até o atendimento ao cliente (MCKINSEY, 2021).

No entanto, é importante notar que essa transformação também trouxe desafios significativos. O estudo de 2021 estima que até 25% da força de trabalho global terá que se adaptar e mudar de ocupação até 2030, o que equivale a aproximadamente 670 milhões de trabalhadores. Além disso, até 30% dos empregos globais, o que representa cerca de 800 milhões de trabalhadores, podem ser completamente eliminados devido à rápida adoção das novas tecnologias (MCKINSEY, 2021).

Essas mudanças têm implicações profundas na oferta e na demanda de trabalho nos próximos anos. É essencial que governos, empresas e trabalhadores estejam preparados para enfrentar esse cenário em evolução. A formação e o desenvolvimento de habilidades adequadas, bem como a criação de políticas de emprego adaptáveis, serão cruciais para garantir uma transição suave para um futuro do trabalho mais digitalizado e automatizado. A colaboração entre setores público e privado também será fundamental para maximizar os benefícios das novas tecnologias e minimizar os impactos negativos nas economias e na sociedade como um todo (MCKINSEY, 2017).

É evidente que o maior desafio reside em garantir que a força de trabalho esteja devidamente preparada e capacitada para enfrentar a transformação em curso. Isso se torna essencial, visto que em nações incapazes de gerenciar eficazmente essa transição, é possível que se depare com um aumento substancial nas taxas de desemprego e com uma significativa diminuição dos níveis salariais (MCKINSEY, 2017). A transformação em questão representa uma mudança profunda e muitas vezes disruptiva nos setores econômicos e no mercado de trabalho. Elas têm o potencial de redefinir os tipos de empregos disponíveis e as habilidades necessárias para ocupá-los.

Um exemplo marcante que ilustra essa preocupação ocorreu durante a greve dos roteiristas nos Estados Unidos em 2023. Essa greve eclodiu em maio de 2023, perdurando por longos 148 dias (BEDINGFIELD, 2023), foi desencadeada por desacordos entre o sindicato de roteiristas e os estúdios. Além do ponto central desse conflito ser sobre a remuneração e os benefícios, os roteiristas também levantaram a questão sobre o uso da IA, como o ChatGPT, no processo criativo. Argumentaram que a IA deveria ser utilizada apenas como uma ferramenta auxiliar destinada a facilitar a pesquisa e a geração de ideias, e não como uma substituta direta para os roteiristas humanos (CULLINS; KILKENNY, 2023).

Como parte do acordo para encerrar a greve, ficou estabelecido que os estúdios e sindicatos de roteiristas irão realizar reuniões a cada semestre para discutir como a IA pode ser integrada ao processo criativo e de produção de filmes. Embora os estúdios não estejam proibidos de usar IA, os escritores têm o direito de processar se seu trabalho for usado para treinar um sistema de IA. Além disso, os

roteiristas podem optar por usar IA em seus trabalhos, mas as empresas não podem forçá-los a fazê-lo. Para garantir uma maior transparência, os estúdios devem informar os roteiristas se o material foi criado por meio de um processo de IA (RICHWINE; CHMIELEWSKI, 2023).

Dessa forma, o acordo busca encontrar um equilíbrio entre o uso da IA e os direitos dos roteiristas, garantindo que sua contribuição criativa seja devidamente reconhecida e protegida (RICHWINE; CHMIELEWSKI, 2023). Em última análise, o sucesso do acordo pode apontar o caminho para um futuro mais equitativo no contexto do trabalho digital. A colaboração entre os diferentes atores envolvidos desempenha um papel fundamental na criação de diretrizes e regulamentações que garantam que a tecnologia seja usada em benefício da humanidade e que os benefícios sejam distribuídos de forma justa e equitativa.

Portanto, a adaptação da força de trabalho a essa nova realidade e o estabelecimento de um diálogo eficaz são fundamentais para promover o bem-estar econômico e social. Isso implica a implementação de uma ampla gama de medidas, como programas de treinamento e requalificação, educação contínua e acesso facilitado a recursos que capacitam os trabalhadores a adquirir as competências necessárias para permanecerem relevantes em um mercado de trabalho em constante evolução. Além disso, é de suma importância que exista uma coordenação efetiva entre o governo, as empresas e as instituições educacionais, com o objetivo de assegurar que as pessoas tenham acesso equitativo às oportunidades de desenvolvimento de habilidades (MCKINSEY, 2017).

Por outro lado, se essa transição não for gerenciada adequadamente, as consequências podem ser severas. O aumento do desemprego pode levar a uma pressão econômica crescente, bem como a um aumento das disparidades sociais. A redução dos salários pode resultar em padrões de vida mais baixos para a população em geral, com potenciais impactos negativos sobre o consumo e o crescimento econômico (MCKINSEY, 2017).

A interação complexa de vários fatores, como o nível salarial, o crescimento da demanda econômica, a demografia e a composição dos setores econômicos e ocupações em cada país, desempenhará um papel fundamental na maneira como a

IA e a automação impactarão diferentes nações (MCKINSEY, 2017). Vamos analisar essa questão com um foco maior no Brasil, que fornece um exemplo ilustrativo desse fenômeno.

No contexto brasileiro, estima-se que até 2030, a adoção moderada das tecnologias mencionadas no artigo possa levar à eliminação potencial de 15,7 milhões de empregos em tempo integral. Essa perspectiva é alarmante, especialmente quando consideramos os dados mais recentes da Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). No segundo trimestre de 2023, a taxa de desemprego atingiu 8%, o que significa que cerca de 8,6 milhões de brasileiros estavam desempregados naquele momento (MCKINSEY, 2017).

A presente conjuntura traz consigo um desafio de proporções consideráveis. Há um potencial evidente para agravar de forma significativa as disparidades de renda já existentes, o que, por sua vez, poderia aprofundar ainda mais as desigualdades sociais e, conseqüentemente, acentuar os desafios estruturais relacionados ao mercado de trabalho brasileiro.

Essa situação configura um desafio de grande magnitude para o país, uma vez que a ausência de crescimento econômico, inovação insuficiente e investimentos inadequados podem intensificar ainda mais a crise no mercado de trabalho. O risco iminente é que, na ausência de medidas adequadas, essa crise possa se aprofundar, potencialmente dobrando os problemas estruturais de emprego que o Brasil já enfrenta atualmente.

A pesquisa destaca a importância de reintegrar rapidamente os trabalhadores deslocados no mercado de trabalho para evitar um aumento ainda maior do desemprego. Esse processo requer não apenas medidas de capacitação e reciclagem profissional, mas também a criação de oportunidades em setores emergentes e o estímulo à inovação e ao empreendedorismo (MCKINSEY, 2017).

Além disso, políticas públicas eficazes e parcerias entre o governo, a indústria e a academia são essenciais para criar um ambiente propício ao crescimento econômico sustentável, à criação de empregos e à adaptação à rápida evolução tecnológica. Investimentos em educação de qualidade, infraestrutura digital e apoio

às pequenas e médias empresas podem desempenhar um papel fundamental nesse processo (MCKINSEY, 2017).

Em 2023, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) lançou um estudo abrangente que analisa minuciosamente o impacto iminente da IA no mercado de trabalho. Esse relatório, baseado em investigações e análises conduzidas pela organização, reforça a preocupação substancial anteriormente apontada pela McKinsey (2017, 2021): aproximadamente 27% dos empregos nos países membros da OCDE enfrentam um risco considerável de automação por meio da IA. Este dado, por si só, ressalta de forma inquestionável a crescente influência da tecnologia na configuração do mercado de trabalho e destaca a urgência de adaptação por parte dos trabalhadores e das políticas públicas.

No referido documento, a organização destaca a necessidade imediata de ação por parte dos países diante dos desafios impostos pela IA ao mercado de trabalho. É enfatizada a importância de políticas proativas para enfrentar esses riscos, sublinhando o papel essencial dos governos nesse processo, uma vez que devem garantir que a IA seja usada como uma ferramenta para promover a criação de mercados de trabalho inclusivos, em vez de prejudicá-los (OCDE, 2023).

Conforme indicado no documento, a influência da IA sobre os trabalhadores e o ambiente de trabalho, bem como a questão de se os benefícios superarem os riscos, estão intrinsecamente ligados às políticas que optamos por adotar. É fundamental reconhecer que a progressão da IA no âmbito profissional deve ser incentivada, uma vez que traz consigo uma série de vantagens e oportunidades significativas. No entanto, é igualmente imperativo que não caiamos na armadilha do "determinismo tecnológico", onde a tecnologia supostamente dita as transformações sociais e culturais, em vez de ser moldada e influenciada por elas (OCDE, 2023).

A OCDE (2023) também chama a atenção para a iminente revolução na dinâmica do mercado de trabalho impulsionada pela expansão da IA. Nesse contexto, é essencial entender quais empregos passarão por transformações, quais serão criados e quais podem desaparecer. Além disso, é fundamental compreender

como as habilidades exigidas dos trabalhadores estão evoluindo nesse cenário de mudança tecnológica.

O documento ainda destaca que a IA pode melhorar significativamente a eficiência, a precisão e a automação de tarefas complexas, permitindo que os trabalhadores se concentrem em atividades mais criativas e estratégicas. No entanto, é crucial abordar com cuidado a implementação da IA e garantir que as políticas sejam formuladas para proteger os direitos dos trabalhadores e minimizar o risco de desigualdades crescentes. Isso inclui o desenvolvimento de programas de requalificação para ajudar os trabalhadores a adquirir habilidades complementares à IA, bem como a criação de regulamentações que assegurem a transparência, a ética e a responsabilidade no uso da tecnologia (OCDE, 2023).

É crucial que a sociedade como um todo participe ativamente do processo de desenvolvimento e implementação da IA. As políticas relacionadas devem ser moldadas com base em valores humanos, éticos e culturais, indo além das considerações puramente tecnológicas. A colaboração entre governo, indústria, academia e sociedade civil desempenha um papel fundamental na garantia de que a IA seja uma ferramenta que aprimore nossa qualidade de vida e capacidades, em vez de limitá-las ou distorcê-las. Portanto, adotar uma abordagem responsável e equilibrada para a IA é essencial na construção de um futuro de trabalho e um ambiente que beneficie a todos (OCDE, 2023).

O documento enfatiza a necessidade de ação imediata para garantir o uso responsável e confiável da IA. Por um lado, é crucial permitir que trabalhadores e empregadores aproveitem os benefícios da IA, ao mesmo tempo em que se adaptam a essa tecnologia, principalmente por meio de programas de treinamento e diálogo social (OCDE, 2023). Além disso, destaca a importância de os países agirem prontamente para enfrentar os impactos da IA no mercado de trabalho, ressaltando a necessidade de políticas que promovam a capacitação, o treinamento e a educação como componentes essenciais dessa abordagem. À medida que a IA continua a moldar o futuro do trabalho, a colaboração entre governos, empregadores e instituições de ensino se torna ainda mais crucial para garantir uma transição suave e justa para todos os membros da sociedade (OCDE, 2023).

Isso implica em uma necessidade de criar políticas públicas inovadoras e fomentar parcerias estratégicas entre os setores público e privado se faz premente. Essa abordagem se revela indispensável para não apenas aproveitar plenamente os benefícios da automação, mas também para mitigar e minimizar os impactos negativos que podem surgir no que tange ao desemprego e às desigualdades sociais (BRYNJOLFSSON; MCAFEE, 2014).

A transformação do mercado de trabalho impulsionada pela IA é inevitável, mas sua implementação deve ser guiada por valores humanos e éticos, visando melhorar a qualidade de vida e as capacidades das pessoas. Nesse sentido, a ação imediata e coordenada se torna fundamental para garantir que a sociedade esteja preparada não apenas para enfrentar os desafios, mas também para colher os benefícios dessa revolução tecnológica em curso (FLORIDI, 2014).

A TAR nos ajuda a compreender que a IA está se tornando um ator central na mediação das formas de trabalho. Emergindo como um ator dominante moldando as relações de trabalho. No entanto, devemos questionar se a IA, frequentemente desenvolvida e controlada por poucas empresas e instituições, está servindo verdadeiramente ao interesse público ou perpetuando desigualdades.

A crescente influência da IA nas dinâmicas de trabalho contemporâneas apresenta uma série de desafios complexos e frequentemente não tão óbvios. Um desses desafios significativos é a automação de ocupações tradicionais, uma tendência que pode resultar em uma crise de desemprego em larga escala e aprofundar ainda mais as disparidades existentes entre aqueles que possuem habilidades pertinentes à IA e aqueles que não as possuem (FORD, 2016).

A requalificação da força de trabalho é frequentemente apontada como a solução ideal para enfrentar os desafios decorrentes das mudanças tecnológicas e econômicas. No entanto, é importante reconhecer que sua eficácia enfrenta questionamentos legítimos, dado o contexto de desigualdades socioeconômicas profundamente enraizadas em nossa sociedade. A ideia de que todos podem simplesmente adquirir novas competências para se manterem relevantes no mercado de trabalho ignora deliberadamente as barreiras econômicas, geográficas e culturais que muitos indivíduos enfrentam.

Para enfrentar esse desafio de maneira eficaz, é fundamental adotar uma abordagem abrangente e equitativa. Isso significa que a requalificação da força de trabalho deve ser complementada por políticas sociais que abordem as desigualdades subjacentes em nossa sociedade. É imperativo garantir que todos tenham igualdade de acesso a oportunidades educacionais e de desenvolvimento, independentemente de sua origem socioeconômica, local de residência ou identidade cultural.

Por isso, é importante reconhecer que a consideração das diferentes agendas dos diversos atores envolvidos é um aspecto crítico na formulação de políticas sociais e econômicas eficazes. No entanto, essa complexidade também pode ser um desafio significativo. E a TAR nos ajuda a entender como diferentes atores, sejam eles indivíduos, organizações ou tecnologias, estão interconectados em uma rede complexa de relações. No contexto do futuro do trabalho, essa abordagem nos permite examinar como as tecnologias emergentes, como a automação e a inteligência artificial, se tornam atores influentes na moldagem das agendas.

A TAR oferece uma lente intrigante para analisar as diferentes agendas que empresas, instituições, governos e pessoas podem perseguir em nossa sociedade complexa e interconectada, especialmente à luz das questões sobre o futuro do trabalho. Essa teoria destaca como os atores, sejam eles humanos ou não, estão envolvidos em redes de interesses, poder e influência, muitas vezes em constante negociação e conflito (AKRICH, 2014). No entanto, essa perspectiva também nos permite identificar algumas críticas significativas em relação às agendas em jogo.

As tecnologias desempenham um papel político nas mudanças da sociedade (AKRICH, 2014). Na rede da IA e principalmente do trabalho, elas podem ser vistas como atores que impulsionam a automação de tarefas, redefinem as qualificações necessárias e reconfiguram as relações de emprego. Sob a perspectiva da TAR, as tecnologias não são apenas ferramentas neutras (LATOURET, 2012), mas sim atores ativos que têm interesses e objetivos próprios, frequentemente associados à eficiência e à maximização de lucros (AKRICH, 2014).

Isso levanta questões cruciais sobre como as agendas tecnológicas são definidas no contexto do trabalho. As empresas, em busca de maior eficiência,

muitas vezes promovem a automação e a robotização, o que pode resultar na substituição de trabalhadores por máquinas (BRYNJOLFSSON; MCAFEE, 2014) (FORD, 2016). Nesse cenário, a TAR nos ajuda a analisar como as empresas colaboram com desenvolvedores de tecnologia e outros atores para implementar essas mudanças, frequentemente sem considerar devidamente o impacto social e econômico sobre os trabalhadores.

Além disso, a TAR nos alerta para o papel das instituições governamentais na formação das agendas do futuro do trabalho. Governos podem agir como mediadores ou reguladores dessas redes de atores, influenciando como as tecnologias são adotadas e quais políticas de proteção social são implementadas para mitigar os impactos negativos das mudanças no mercado de trabalho.

No entanto, a TAR também levanta questões sobre o poder assimétrico nas redes. Grandes corporações e empresas de tecnologia muitas vezes têm uma vantagem significativa na formação das agendas devido aos seus recursos financeiros e influência política. Isso pode resultar em agendas tecnológicas que priorizam os interesses das empresas em detrimento dos interesses dos trabalhadores e da sociedade em geral (ZUBOFF, 2019). No contexto do futuro do trabalho, isso pode significar que as agendas das empresas, como a automação e a otimização de custos, podem ser promovidas sem considerar adequadamente o impacto humano, levantando questões sobre o bem-estar dos trabalhadores (BRYNJOLFSSON; MCAFEE, 2014).

Outra crítica relevante sobre a questão do poder assimétrico nas redes de atores, é o que se torna ainda mais pronunciado quando se trata do futuro do trabalho. Nem todos os atores têm a mesma capacidade de moldar agendas ou influenciar decisões (BACHRACH; BARATZ, 1962). Isso pode resultar em desequilíbrios significativos de poder, onde empresas e instituições, atores, com recursos financeiros consideráveis têm uma vantagem injusta na definição das políticas de trabalho, enquanto os trabalhadores, também atores, podem se encontrar em posições mais vulneráveis.

Além disso, pode nos levar a entender que todas as agendas são igualmente válidas, desde que sejam "negociadas" dentro das redes de interesse. No contexto

do futuro do trabalho, isso levanta preocupações sobre como as agendas de empresas que buscam maximizar lucros podem influenciar a adoção de tecnologias que resultam em desemprego ou precarização do trabalho (AUTOR, 2015). É essencial considerar as implicações éticas e sociais dessas agendas.

Em última análise, a Teoria Ator-Rede oferece uma metodologia e perspectiva valiosas para compreender como as agendas são moldadas e negociadas em um mundo cada vez mais complexo e interconectado, abrangendo preocupações que vão desde o futuro do trabalho até questões sociotecnológicas mais amplas. No entanto, é imperativo buscar um equilíbrio cuidadoso. Enquanto reconhecemos a intrincada teia de interações entre atores humanos e não humanos, também devemos garantir que as decisões tomadas dentro dessas redes sejam conduzidas pelo imperativo do bem-estar, justiça e equidade em nossa sociedade.

Isso implica uma abordagem responsável ao desenvolvimento e implementação de tecnologias, bem como a formulação de políticas que considerem os impactos sociais, éticos e econômicos em todas as fases do processo. A TAR nos lembra que os atores das redes sociotécnicas são dinâmicos e interdependentes, e as decisões tomadas em um ponto podem reverberar em toda a rede (LATOUR, 2012). Portanto, ao explorar o potencial da TAR para a análise crítica das complexas questões contemporâneas, também devemos estar comprometidos em garantir que nossas ações e políticas promovam uma sociedade mais justa e inclusiva. É nessa interseção entre compreensão complexa e responsabilidade ética que encontramos o caminho para um futuro mais equitativo e sustentável.

6 CONCLUSÕES

A IA desempenha um papel fundamental na transformação de diversos aspectos da sociedade contemporânea. No entanto, é essencial reconhecer que seu avanço traz consigo uma série de desafios éticos, sociais e práticos que exigem nossa atenção e abordagem cuidadosa. A abordagem da Teoria da Ator-Rede (TAR) surge como uma ferramenta valiosa para compreender a IA como uma construção social intrincada e interdependente, uma rede sociotécnica. Isso nos permite enxergar a IA não como uma entidade isolada, mas como parte de uma rede

complexa de atores que inclui seres humanos, organizações e tecnologias. Dentro dessa rede, esses atores interagem de maneiras diversas e mutuamente influentes, desempenhando papéis cruciais no processo de desenvolvimento e uso da IA.

A IA é moldada e influenciada por uma série de fatores sociais, que abrangem desde crenças culturais até interesses econômicos, agendas políticas e as dinâmicas complexas de poder presentes na sociedade. Essa visão nos permite compreender que a IA não é uma criação puramente técnica, mas um produto da interação humana e social.

Esta dissertação buscou realizar uma análise crítica da IA, revelando uma série de aspectos sociais que merecem atenção e investigação aprofundada. Sob a perspectiva da TAR, podemos abordar a complexa rede de atores e interações que moldam o desenvolvimento e a evolução da IA.

No contexto da rápida disseminação da IA, surgem uma série de questões que merecem nossa atenção. Mas, a questão central diz respeito à definição dos parâmetros que devem guiar o desenvolvimento dessa tecnologia. Quem, afinal, deve ser responsável por determinar os valores e princípios que nortearão a IA? É crucial garantir que essas decisões sejam tomadas de maneira transparente e inclusiva, com a participação de diversas partes interessadas.

As questões éticas, de privacidade e de segurança também representam desafios significativos para a capacidade da comunidade de lidar de forma responsável com a IA. A TAR nos lembra que, por trás das inovações e promessas de transformação, existem complexas dinâmicas de poder e interesses em jogo. Também levanta preocupações substanciais sobre a possibilidade de a IA ser moldada por interesses que podem não estar necessariamente alinhados com o bem-estar da sociedade. Portanto, é imperativo questionar não apenas quem define os parâmetros da IA, mas também quem determina quais dados serão usados para treiná-la e quem terá a autoridade para tomar decisões sobre sua aplicação.

Isso nos leva a compreender a IA como uma rede interconectada de elementos, o que, por sua vez, gera implicações cruciais quando se trata de atribuir responsabilidades em cenários que envolvam problemas éticos ou de segurança. Essa complexidade de atores pode levar a desafios substanciais na determinação

de responsabilidades claras em casos de problemas ou falhas. A importância crítica de considerar as implicações éticas e sociais da IA é um questionamento importante.

A IA pode contribuir para a melhoria da qualidade de vida e eficiência em vários setores, mas também pode levantar questões sérias sobre privacidade, discriminação algorítmica, concentração de poder e o desemprego. Portanto, compreender a IA como uma construção social nos leva a questionar quem se beneficia com essa tecnologia e quem é prejudicado, e como podemos garantir que ela seja desenvolvida e usada de maneira ética e justa.

Um ponto crítico a ser destacado é o papel significativo das grandes empresas de tecnologia, como atores, que podem influenciar a formulação de políticas e regulamentações governamentais relacionadas à IA. A influência desproporcional dessas empresas levanta sérias preocupações sobre a possibilidade de conflitos de interesses, nos quais os objetivos dessas empresas podem prevalecer em detrimento do bem-estar público.

É de extrema importância compreender profundamente a problemática subjacente a essa questão, sobretudo quando os interesses das corporações econômicas exercem uma influência desmedida sobre as decisões, muitas vezes relegando as considerações éticas e sociais a segundo plano. Assim, torna-se imprescindível trabalhar na busca por uma garantia de que as escolhas e determinações que emanam desse contexto estejam em consonância com os valores humanos, éticos e culturais que moldam nossa sociedade.

O reconhecimento da complexidade das redes de atores é fundamental para se ter uma visão completa e precisa do quadro geral. Nesse emaranhado de relacionamentos, influências e interesses, é fácil perder de vista as implicações éticas e sociais de nossas ações. No entanto, ignorar essas considerações pode levar a consequências graves, principalmente a desigualdade crescente e a deterioração dos valores culturais que sustentam nossa comunidade global.

A busca pelo equilíbrio é um imperativo moral e prático. Devemos assegurar que as decisões tomadas no âmbito empresarial e político sejam guiadas por princípios que respeitem a dignidade humana, promovam a justiça social e preservem a diversidade cultural. Isso requer um esforço conjunto de todas as

partes interessadas, desde governos e empresas até organizações da sociedade civil e cidadãos comuns.

Outro ponto crítico é o poder de influência que a IA pode exercer sobre as vidas das pessoas. Aqueles que controlam os algoritmos e os dados subjacentes têm um poder considerável sobre o comportamento e as escolhas das pessoas. Portanto, é necessário questionar quem são os principais atores por trás do desenvolvimento e implementação da IA e garantir que haja uma representação diversificada de vozes e perspectivas para evitar a concentração excessiva de poder.

A discriminação algorítmica é outra preocupação na aplicação da IA, uma vez que os algoritmos podem inadvertidamente refletir preconceitos existentes na sociedade. Para evitar isso, é fundamental implementar regulamentações e supervisões rigorosas que garantam que os sistemas de IA sejam justos e imparciais. Isso implica a criação de modelos de treinamento que sejam sensíveis a questões de equidade, a coleta de dados diversificados e a realização de auditorias regulares para identificar e corrigir viés nos algoritmos. Além disso, a escolha dos dados utilizados para treinar os modelos de IA é de extrema importância, já que isso pode ter um profundo impacto na qualidade e imparcialidade dos resultados gerados.

Outras complexidades e desafios significativos estão emergindo na sociedade contemporânea quanto à adaptação e transformação do mercado de trabalho impulsionada pela IA. As preocupações com desemprego em massa, desigualdade econômica e automação de empregos tradicionais apresentam estimativas alarmantes sobre o potencial impacto da IA no emprego. As implicações desses desafios para a sociedade são profundas, tornando crucial a discussão sobre a requalificação da força de trabalho e a implementação de políticas sociais que abordem as desigualdades subjacentes. Sob a perspectiva da TAR também podemos enfatizar o papel das empresas na definição das agendas tecnológicas e como o poder assimétrico nas redes de atores pode influenciar decisões que, em última instância, podem prejudicar os trabalhadores.

Buscamos enfatizar a importância de uma abordagem crítica inspirada pela TAR ao lidar com a crescente integração da IA em nossa sociedade. É importante que a comunidade de IA esteja atenta às questões levantadas e trabalhe de forma colaborativa para garantir que a tecnologia seja desenvolvida e aplicada de maneira responsável, assegurando que seus benefícios sejam distribuídos de maneira justa e equitativa em toda a sociedade.

Além disso, nossa pesquisa buscou demonstrar que a IA não é uma entidade estática, mas sim um campo em constante evolução. Os avanços tecnológicos e as mudanças nas dinâmicas sociais estão moldando continuamente a IA e sua interação com a sociedade. A rápida expansão da IA levanta questões sobre regulamentação, governança e responsabilidade. Como uma construção social, a IA requer uma abordagem holística que envolva não apenas cientistas e engenheiros, mas também governos, organizações da sociedade civil e a população em geral.

Em última análise, esta conclusão enfatiza que a IA não pode ser vista como uma ferramenta neutra, mas sim como um reflexo dos valores, intenções e influências de seus criadores e controladores. Ela é uma manifestação das escolhas éticas, culturais e sociais que permeiam sua concepção e desenvolvimento. Portanto, é de importância que adotemos uma abordagem continuamente crítica e reflexiva à medida que a IA se torna cada vez mais uma parte intrínseca de nossa sociedade.

A IA não é apenas uma tecnologia emergente; é uma força que moldará (ou já está moldando) muitos aspectos de nossas vidas, desde o trabalho e a economia até a saúde e a educação. Portanto, é fundamental considerá-la como uma questão central na sociedade contemporânea. Devemos estar dispostos a realizar uma análise constante e a fazer ajustes à medida que novos desafios e dilemas éticos relacionados à IA emergem.

Um dos principais imperativos é assegurar que o poder e os benefícios da IA sejam distribuídos de maneira justa e equitativa. Isso significa que a sociedade deve se esforçar para evitar que a IA seja monopolizada por interesses seletos em detrimento do bem comum. Em vez disso, devemos buscar políticas e regulamentações que promovam a inclusão, a diversidade e a acessibilidade na IA.

Isso inclui a consideração atenta das questões de viés algorítmico e discriminação, bem como a garantia de que grupos marginalizados não sejam prejudicados ou excluídos pelas aplicações de IA. Além de questionar os valores subjacentes aos algoritmos, promover a transparência em relação às decisões algorítmicas e buscar maneiras de mitigar os impactos negativos da IA, como o impacto nos empregos.

Ressaltando a importância de uma abordagem interdisciplinar para entender a IA como uma construção social. A colaboração entre especialistas em tecnologia, ética, sociologia, política e outras disciplinas é fundamental para abordar os desafios complexos que a IA apresenta. É por meio desse diálogo e cooperação que podemos desenvolver políticas e práticas que promovam o uso ético e justo da IA em nossa sociedade.

Em suma, a IA é uma força em constante evolução que está redefinindo nossa sociedade em várias frentes. No entanto, essa transformação não ocorre em um vácuo; ela é intrinsecamente social e interdependente. A TAR se apresenta como uma valiosa lente para compreender a complexidade dessa construção social e nos lembra que as decisões tomadas e os interesses defendidos têm implicações éticas, sociais e práticas significativas. Portanto, devemos adotar uma abordagem crítica e reflexiva à medida que a IA se integra cada vez mais em nossas vidas. A distribuição justa de poder e benefícios deve ser um objetivo fundamental, com políticas que promovam a inclusão e a diversidade tanto no desenvolvimento da IA quanto nas políticas que envolvam a mesma.

Nossa capacidade de abraçar a IA como uma construção social e tomar medidas responsáveis para orientar seu desenvolvimento moldará o futuro da nossa sociedade de forma significativa. É importante lembrar que a IA é uma área em constante evolução, e a reflexão contínua e a colaboração entre diversas partes interessadas (atores) são essenciais para moldar seu futuro de maneira responsável. Por meio de um compromisso contínuo com a reflexão, a ética e a colaboração, nos permitirá colher os benefícios da IA enquanto mitigamos seus potenciais riscos.

REFERÊNCIAS

AKRICH, M. L'analyse socio-technique. *In*: VINCK, D. **La gestion de la recherche: Nouveaux problèmes, nouveaux outils**. Bruxelles: De Boeck, 1991. par. III. Nouveaux outils, p. 109-134. ISBN 2804114341.

AKRICH, M. The De-Description of Technical Objects. *In*: BIJKER, W. E.; LAW, J. **Shaping Technology/Building Society: studies in sociotechnical change**. 2ª impressão. Cambridge, Massachusetts: The MIT Press, 1997. cap. 7, p. 205-224. ISBN 0-262-02338-5. Disponível em: <https://hdl.handle.net/2027/heb01128.0001.001>. Acesso em: 24 jan. 2023.

AKRICH, M. La construction d'un système socio-technique: Esquisse pour une anthropologie des techniques. *In*: AKRICH, M.; CALLON, M.; LATOUR, B. **Sociologie de la traduction: Textes fondateurs**. Paris: Presses des Mines, 2006. p. 109-134. ISBN 9782356710239. Disponível em: <http://books.openedition.org/pressesmines/1195>. Acesso em: 24 jan. 2023.

AKRICH, M. Como descrever os objetos técnicos?. **Boletim Campineiro de Geografia**, [s. l.], v. 4, n. 1, p. 161–182, 2014. DOI: 10.54446/bcg.v4i1.147. Disponível em: <https://publicacoes.agb.org.br/boletim-campineiro/article/view/2516>. Acesso em: 24 jan. 2023.

ALZUBAIDI, L. *et al.* Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Journal of Big Data**, [s. l.], v. 8, n. 53, 31 mar. 2021. DOI 10.1186/s40537-021-00444-8. Disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8>. Acesso em: 16 maio 2023.

AMORIM, C. **A Máquina e Seus Limites: Uma Investigação Sobre o Xadrez Computacional**. Orientador: Robinson Moreira Tenório. 2002. 142 f. Dissertação (Mestrado em Ensino, Filosofia e História das Ciências) - Universidade Federal da Bahia e Universidade Estadual de Feira de Santana, Salvador, 2002. Disponível em: <https://ppgefhc.ufba.br/pt-br/maquina-e-seus-limites-uma-investigacao-sobre-o-xadrez-computacional>. Acesso em: 7 maio 2023.

AUMÜLLER-WAGNER, S. Encoded Bias in Recruitment Algorithms: A practical, intersectional approach to fairness in Machine Learning Algorithms. **Excellent Student Paper Series: STS Encounters**, [s. l.], 2019. Disponível em: https://www.dasts.dk/wp-content/uploads/Sophia-Aumueller-Wagner_2019.pdf. Acesso em: 16 jul. 2023.

AUTOR, D. H. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. **Journal of Economic Perspectives**, [s. l.], v. 29, n. 3, p. 3-30, verão 2015. DOI 10.1257/jep.29.3.3. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.29.3.3>. Acesso em: 19 set. 2023.

BACHRACH, P.; BARATZ, M. S. Two Faces of Power. **The American Political Science Review**, [s. l.], v. 56, n. 4, p. 947-952, Dez. 1962. DOI 10.2307/1952796. Disponível em: <https://www.jstor.org/stable/1952796>. Acesso em: 19 set. 2023.

BARBOSA, X. C.; BEZERRA, R. F. BREVE INTRODUÇÃO À HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL. **Jamaxi**, [s. l.], v. 4, n. 1, 2020. Disponível em: <https://periodicos.ufac.br/index.php/jamaxi/article/view/4730>. Acesso em: 16 abr. 2023.

BEDINGFIELD, W. Hollywood Writers Reached an AI Deal That Will Rewrite History. **WIRED**, [s. l.], 26 set. 2023. Disponível em: <https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/>. Acesso em: 9 out. 2023.

BESSEN, J. AI and Jobs: the role of demand. **NBER Working Papers**, Cambridge, MA, n. 24235, jan. 2018. DOI 10.3386/w24235. Disponível em: <https://www.nber.org/papers/w24235>. Acesso em: 19 set. 2023.

BLEAKLEY, C. **Poems That Solve Puzzles: The History and Science of Algorithms**. 1ª. ed. Oxford: Oxford University Press, 2020. 288 p. ISBN 9780198853732.

BOUCHER, P. Artificial intelligence: How does it work, why does it matter, and what can we do about it?. *In*: EUROPEAN PARLIAMENTARY RESEARCH SERVICE (org.). **Panel for the Future of Science and Technology**. [s. l.], jun. 2020. Disponível em: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2020\)641547](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641547). Acesso em: 17 maio 2023.

BUOLAMWINI, J.; GEBRU, T. Gender Shades. *Proceedings of Machine Learning Research: Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, [s. l.], n. 81, p. 77-91, 2018. Conference on Fairness, Accountability and Transparency, 23-24 fev. 2018, Nova Iorque, NY, EUA. Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a.html>. Acesso em: 10 maio 2023.

BRYNJOLFSSON, E.; MCAFEE, A. **The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies**. Nova Iorque: W. W. Norton & Company, 2014. 336 p. ISBN 9780393241259. *Kindle Edition*.

CALLON, M. The state and technical innovation: a case study of the electrical vehicle in France. **Research Policy**, [s. l.], v. 9, ed. 4, p. 358-376, out. 1980. DOI 10.1016/0048-7333(80)90032-3. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0048733380900323>. Acesso em: 22 fev. 2023.

CALLON, M. Some elements of a sociology of translation: domestication of the scallops and the fisherman of St. Briec Bay. *In*: LAW, J. (ed.). **Power, action, and belief: a new sociology of knowledge?**. Boston: Routledge & Kegan Paul, 1986. p. 196-223. ISBN 0710208022.

CAMILLIS, P. K.; BUSSULAR, C. Z.; ANTONELLO, C. S. A agência a partir da Teoria Ator-Rede: reflexões e contribuições para as pesquisas em administração.

Organizations & Society, Salvador, v. 23, n. 76, p. 73-91, Jan-Mar 2016. DOI 10.1590/1984-9230764. Disponível em:

<https://www.scielo.br/j/osoc/a/mk7zPp8vV9fGr6fjJtndmD/?lang=pt>. Acesso em: 17 set. 2023.

CAO, Y. et al. A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. **ArXiv (preprint)**, *online*, 7 mar. 2023. DOI

10.48550/arXiv.2303.04226. Disponível em: <https://arxiv.org/abs/2303.04226>. Acesso em: 20 set. 2023.

CARDOSO, T.; SANTAELLA, L. A relevância da mediação no pensamento de Bruno Latour. *In*: ALZAMORA, G.; ZILLER, J.; COUTINHO, F. (org.). **Dossiê Bruno Latour**. Belo Horizonte: Editora UFMG, 2020. par. II - Mediação e Hibridismo, p. 141-178. ISBN 9788542303223.

CATALANO, J. V. C.; LORENZI, B. R. Sem Referências: o ChatGPT sob a perspectiva latouriana e a armadilha do Duplo Clique. **Revista Faz Ciência**, [s. l.], v. 25, n. 41, 2023. DOI: 10.48075/rfc.v25i41.30761. Disponível em:

<https://e-revista.unioeste.br/index.php/fazciencia/article/view/30761>. Acesso em: 20 set. 2023.

CHANG, E. **Brotopia**: Breaking Up The Boys' Club Of Silicon Valley. 1ª. ed. Nova Iorque: Portfolio/Penguin, 2018. 336 p. ISBN 9780525540175.

CHATTERJEE, J.; DETHLEFS, N. This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. **Patterns**, [s. l.], v. 4, n. 100676, ed. 1, 13 jan. 2023. DOI 10.1016/j.patter.2022.100676. Disponível em:

[https://www.cell.com/patterns/fulltext/S2666-3899\(22\)00323-3](https://www.cell.com/patterns/fulltext/S2666-3899(22)00323-3). Acesso em: 14 fev. 2023.

CHEN, X. *et al.* How Robust is GPT-3.5 to Predecessors?: A Comprehensive Study on Language Understanding Tasks. **ArXiv (preprint)**, *online*, 1 mar. 2023. DOI

10.48550/arXiv.2303.00293. Disponível em: <https://arxiv.org/abs/2303.00293>. Acesso em: 14 maio 2023.

CHUI, M.; MANYIKA, J.; MIREMADI, M. Where machines could replace humans—and where they can't (yet). **McKinsey Quarterly**, [s. l.], 8 jul. 2016.

Disponível em:

<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet>. Acesso em: 19 set. 2023.

COECKELBERGH, M. **AI Ethics**. 1ª. ed. Cambridge, Massachusetts: The MIT Press, 2020. 248 p. ISBN 9780262538190.

CORRÊA, D. S. Do problema do social ao social como problema: elementos para uma leitura da sociologia pragmática francesa. **REVISTA DE CIÊNCIAS SOCIAIS - POLÍTICA & TRABALHO**, [s. l.], v. 1, n. 40, 2014. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/politicaetrabalho/article/view/18140>. Acesso em: 01 fev. 2023.

CULLINS, A.; KILKENNY, K. As Writers Strike, AI Could Covertly Cross the Picket Line. **The Hollywood Reporter**, [s. l.], 3 maio 2023. Business News. Disponível em: <https://www.hollywoodreporter.com/business/business-news/writers-strike-ai-chatgpt-1235478681/>. Acesso em: 28 set. 2023.

DWIVEDI, Y. K. *et al.* Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. **International Journal of Information Management**, [s. l.], v. 7, n. 102642, Ago. 2023. DOI 10.1016/j.ijinfomgt.2023.102642. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401223000233>. Acesso em: 17 set. 2023.

FEI, N. *et al.* Towards artificial general intelligence via a multimodal foundation model. **Nature Communications**, [s. l.], ano 2022, v. 13, n. 3094, 2 jun. 2022. DOI 10.1038/s41467-022-30761-2. Disponível em: <https://www.nature.com/articles/s41467-022-30761-2>. Acesso em: 15 maio 2023.

FLORIDI, L. **The Fourth Revolution: How the Infosphere is Reshaping Human Reality**. 1ª. ed. Inglaterra: Oxford University Press, 2014. 248 p. ISBN 9780199606726.

FORD, M. **Rise of the Robots: Technology and the Threat of a Jobless Future**. Reimpressão. ed. Nova Iorque: Basic Books, 2016. 368 p. ISBN 9780465097531.

FREY, C. B.; OSBORNE, M. A. The future of employment: How susceptible are jobs to computerisation?. **Technological Forecasting and Social Change**, [s. l.], v. 114, p. 254-280, Jan. 2017. DOI 10.1016/j.techfore.2016.08.019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0040162516302244>. Acesso em: 12 set. 2023.

FRISCHMANN, B.; SELINGER, E. **Re-Engineering Humanity**. 1ª. ed. Cambridge, Reino Unido: Cambridge University Press, 2018. 430 p. ISBN 9781107147096.

FUEGI, J.; FRANCIS, J. Lovelace & Babbage and the creation of the 1843 ‘notes’. **IEEE Annals of the History of Computing**, [s. l.], p. 16-26, Oct.-Dec. 2003. DOI 10.1109/MAHC.2003.1253887. Disponível em: <https://ieeexplore.ieee.org/document/1253887>. Acesso em: 7 nov. 2023.

GALLOWAY, Scott. **The Four: The Hidden DNA of Amazon, Apple, Facebook, and Google**. 1ª. ed. Nova Iorque: Portfolio/Penguin, 2017. 323 p. ISBN 9780735213654.

GARTNER. What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle. **Gartner Insights**, *online*, 15 set. 2022. Disponível em: <https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle>. Acesso em: 15 maio 2023.

GOERTZEL, T. The Path to More General Artificial Intelligence. **Journal of Experimental & Theoretical Artificial Intelligence**, [s. l.], v. 26, ed. 3, p. 343-354, 8 abr. 2014. DOI 10.1080/0952813X.2014.895106. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/0952813X.2014.895106?journalCode=eta20>. Acesso em: 15 maio 2023.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning** (Adaptive Computation and Machine Learning series). ed. ilustr. Cambridge, Massachusetts: The MIT Press, 2016. 800 p. ISBN 978-0262035613.

HAMMOND, K. **Practical Artificial Intelligence For Dummies®**, Narrative Science Edition. Hoboken, Nova Jersey: John Wiley & Sons, 2015. 52 p. ISBN 9781119149842. *E-book*.

HANSON, R. **The Age of Em: Work, Love, and Life When Robots Rule the Earth**. 1ª. ed. Nova Iorque: Oxford University Press, 2016. v. 528. ISBN 9780198754626.

HURWITZ, J.; KIRSCH, D. **Machine Learning For Dummies®**, IBM Limited Edition. Hoboken, Nova Jersey: John Wiley & Sons, 2018. 74 p. ISBN 9781119454946. *E-book*.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Indicadores IBGE**. Disponível em: <https://www.ibge.gov.br/explica/desemprego.php> Acesso em: 08 set. 2023.

KHAN, S. *et al.* Transformers in Vision: Survey. **ACM Computing Surveys**, [s. l.], v. 54, n. 200, ed. 10s, p. 1–41, 13 set. 2022. DOI 10.1145/3505244. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3505244>. Acesso em: 14 set. 2023.

KIETZMANN, J.; LEE, L. W.; MCCARTHY, I. P.; KIETZMANN, T. C. Deepfakes: Trick or treat?. **Business Horizons**, [s. l.], v. 63, n. 2, p. 135-146, 12 fev. 2020. DOI 10.1016/j.bushor.2019.11.006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0007681319301600>. Acesso em: 27 set. 2023.

KURZWEIL, R. **The Singularity Is Near: When Humans Transcend Biology**. 1ª. ed. Londres: Duckworth Overlook, 2005. 652 p. ISBN 9780715639382. *E-book*.

LATOUR, B. **The Pasteurization of France**. Tradução: Alan Sheridan, John Law. Cambridge, Massachusetts: Harvard University Press, 1988. 288 p. ISBN 9780674657601.

LATOUR, B. On Technical Mediation. **Common Knowledge**, [s. l.], v. 3, ed. 2, p. 29-64, outono 1994. Disponível em: <http://www.bruno-latour.fr/node/234.html>. Acesso em: 23 fev. 2023.

LATOUR, B. ; WOOLGAR, S. **A vida de laboratório: A produção dos fatos científicos**. Tradução: Ângela Ramalho Vianna. 1ª. ed. Rio de Janeiro: Relume Dumará, 1997. 310 p. ISBN 857316123X.

LATOUR, B. **Ciência em ação: como seguir cientistas e engenheiros sociedade afora**. Tradução: Ivone C. Benedetti, Jesus de Paula Assis. São Paulo: Editora Unesp, 2000. 440 p. ISBN 857139265X.

LATOUR, B. **A esperança de Pandora: ensaios sobre a realidade dos estudos científicos**. Tradução: Gilson César Cardoso de Sousa. 1ª. ed. Bauru: EDUSC, 2001. 372 p. ISBN 8574600628.

LATOUR, B. **Reagregando o social: uma introdução à Teoria do Ator-Rede**. Tradução: Gilson César Cardoso de Sousa. 1ª. ed. Salvador: EDUFBA; Bauru: EDUSC, 2012. 400 p. ISBN 9788523208646.

LATOUR, B. **Jamais fomos modernos: Ensaio de antropologia simétrica**. Tradução: Carlos Irineu da Costa. 3ª. ed. São Paulo: Editora 34, 2013. 192 p. ISBN 9788573267396.

LAW, J. On the methods of long-distance control – vessels, navigation and the portuguese route to India. **The Sociological Review**, [s. l.], v. 32, n. 1, ed. sup., p. 234-263, maio 1984. DOI 10.1111/j.1467-954X.1984.tb00114.x. Disponível em: <https://journals.sagepub.com/doi/10.1111/j.1467-954X.1984.tb00114.x>. Acesso em: 9 fev. 2023.

LEE, K. **AI Superpowers: China, Silicon Valley, and the New World Order**. 1ª. ed. Boston: Houghton Mifflin Harcourt Publishing Company, 2018. 272 p. ISBN 9781328546395. *Kindle Edition*.

LI, F. **I often tell my students not to be misled by the name ‘artificial intelligence’ [...]** 13 fev. 2018. X (Twitter): @drfeifei. Disponível em: <https://twitter.com/drfeifei/status/963564896225918976>. Acesso em: 24 set. 2023.

LIGHTHILL, J. **Artificial Intelligence: A General Survey**, *online*, jul. 1972. Disponível em: <https://www.aii.ed.ac.uk/events/lighthill1973/lighthill.pdf>. Acesso em: 30 abr. 2023.

LOWE, S. D. **Deep Learning For Dummies®**, HPE Special Edition. Hoboken, Nova Jersey: John Wiley & Sons, 2018. 52 p. ISBN 9781119484530. *E-book*.

LUGER, G. F. **Inteligência artificial**. Tradução: Daniel Vieira. 6ª. ed. São Paulo: Pearson, 2013. 632 p. ISBN 9788581435503.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 9ª. ed. São Paulo: Atlas, 2021. 368 p. ISBN 9788597026566.

MCKINSEY & COMPANY. **O futuro do mercado de trabalho**: impacto em empregos, habilidades e salários. Dezembro de 2017. Disponível em: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages/ptbr> Acesso em: 16 mar. 2023.

MCKINSEY & COMPANY. **O futuro do trabalho pós-COVID-19**. Fevereiro de 2021. Disponível em: <https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19/pt-br> Acesso em: 31 mar. 2023.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. **A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence**. [s. l.], 31 ago. 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 8 abr. 2023.

MCCORDUCK, P. **Machines Who Think**: A Personal Inquiry Into the History and Prospects of Artificial Intelligence. 2ª. ed. aum. Natick, Massachusetts: A K Peters, 2004. 598 p. ISBN 1568812051

MITCHELL, M. **Artificial Intelligence**: A Guide for Thinking Humans. ed. ilustr. Nova Iorque: Farrar, Straus and Giroux, 2019. 336 p. ISBN 9780374715236. *Kindle Edition*.

MOL, A. Actor-Network Theory: sensitive terms and enduring tensions. **Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft**, Amesterdã, v. 50, p. 253-269, 2010. Disponível em: <https://hdl.handle.net/11245/1.330874>. Acesso em: 7 mar. 2023.

MONDAL, S.; DAS, S.; VRANA, V. G. How to Bell the Cat? A Theoretical Review of Generative Artificial Intelligence towards Digital Disruption in All Walks of Life. **Technologies**, [s. l.], v. 11, n. 2, 17 mar. 2023. DOI 10.3390/technologies11020044. Disponível em: <https://www.mdpi.com/2227-7080/11/2/44>. Acesso em: 18 set. 2023

MOORE, G. E. Cramming More Components Onto Integrated Circuits. **Proceedings of the IEEE**, [s. l.], v. 86, n. 1, p. 82-85, Jan. 1988. DOI 10.1109/JPROC.1998.658762. Disponível em: <https://ieeexplore.ieee.org/document/658762>. Acesso em: 17 set. 2023.

NEWQUIST, H. P. **The Brain Makers**: Genius, Ego, And Greed In The Quest For Machines That Think. 2ª. ed. atual. [s. l.]: The Relayer Group, 2020. 696 p. ISBN 9780988593718. *Kindle Edition*.

NOBLE, S. U. **Algorithms of Oppression**: How Search Engines Reinforce Racism. 1ª. ed. Nova Iorque: NYU Press, 2018. 248 p. ISBN 9781479837243.

O'NEIL, C. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução: Rafael Abraham. 1ª. ed. Santo André: Editora Rua do Sabão, 2020. 342 p. ISBN 9786586460025.

OCDE. **OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market**. Paris: OECD Publishing, 2023. ISBN 9789264427679. DOI 10.1787/08785bba-en. Disponível em: https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2023_08785bba-en. Acesso em: 8 set. 2023.

PALÁCIOS, E. M. G. *et. al.* Introdução aos estudos CTS (Ciência, Tecnologia e Sociedade). **Cadernos de Ibero-América**. Tradução: Irlan von Linsingen, Luiz Teixeira do Vale Pereira, Walter Antonio Bazzo. ed. trad. rev. [s. l.]: Organização dos Estados Ibero-americanos, 2003. 170 p.

PASQUALE, F. **The Black Box Society: The Secret Algorithms That Control Money and Information**. 1ª. ed. Cambridge, Massachusetts: Harvard University Press, 2015. 320 p. ISBN 9780674368279.

RICH, E.; KNIGHT, K.; NAIR, S. B. **Artificial Intelligence**. 3ª. ed. Nova Delhi: Tata McGraw-Hill, 2010. 568 p. ISBN 9780070087705.

RICHWINE, L.; CHMIELEWSKI, D. Hollywood writers guild ends strike ahead of final contract vote. **Reuters**, [s. l.], 27 set. 2023. World at Work. Disponível em: <https://www.reuters.com/world/us/hollywood-writers-guild-calls-end-strike-wednesday-2023-09-27/>. Acesso em: 28 set. 2023.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. Tradução: Regina Célia Simille. 3ª. ed. Rio de Janeiro: Elsevier, 2013. 1016 p. ISBN 9788535237016.

SALGADO, T. P. B. **Sociologias pragmáticas e cultura digital**. 1ª. ed. Salvador: EDUFBA, 2022. 76 p. ISBN 9786556304205. Disponível em: <https://repositorio.ufba.br/handle/ri/36419>. Acesso em: 10 jan. 2023.

SANTAELLA, L.; CARDOSO, T. O desconcertante conceito de mediação técnica em Bruno Latour. **MATRIZES**, [s. l.], v. 9, n. 1, p. 167-185, 2015. DOI: 10.11606/issn.1982-8160.v9i1p167-185. Disponível em: <https://www.revistas.usp.br/matrizes/article/view/100679>. Acesso em: 24 jan. 2023.

SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, [s. l.], v. 2, n.160, p. 1-21, 22 mar. 2021. DOI 10.1007/s42979-021-00592-x. Disponível em: <https://link.springer.com/article/10.1007/s42979-021-00592-x>. Acesso em: 16 maio 2023.

SCHNEIER, B. **Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World**. 1ª. ed. Nova Iorque: W. W. Norton & Company, 2016. 448 p. ISBN 9780393352177.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences**, [s. l.], v. 3, ed. 3, p. 417-424, setembro 1980. DOI <https://doi.org/10.1017/S0140525X00005756>. Disponível em: <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A>. Acesso em: 30 mar. 2023.

SENG, K. P.; ANG, L. M.; NGHARAMIKE, E. Artificial intelligence Internet of Things: A new paradigm of distributed sensor networks. **International Journal of Distributed Sensor Networks**, [s. l.], v. 18, n. 3, 12 mar. 2022. DOI 10.1177/15501477211062835. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/15501477211062835>. Acesso em: 17 set. 2023.

SILVA NETO, V. J.; BONACELLI, M. B. M.; PACHECO, C. A. O sistema tecnológico digital: inteligência artificial, computação em nuvem e Big Data. **Revista Brasileira de Inovação**, Campinas, v. 19, n. e0200024, p. 1-31, 28 dez. 2020. DOI 10.20396/rbi.v19i0.8658756. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rbi/article/view/8658756>. Acesso em: 6 set. 2023.

SILVA NETO, V. J. **The consolidation of the digital technological system: transformations in science and other domains**. Orientador: Maria Beatriz Machado Bonacelli. 2022. 184 p. Tese (Doutorado em Política Científica e Tecnológica) - Universidade Estadual de Campinas, Instituto de Geociências, Campinas, 2022. Disponível em: <https://hdl.handle.net/20.500.12733/6085>. Acesso em: 4 set. 2023.

SISMONDO, S. Science and Technology Studies and an Engaged Program: I. In: HACKETT, E. J.; AMSTERDAMSKA, O.; LYNCH, M.; WAJCMAN, J. (ed.). **The Handbook of Science and Technology Studies**: Third Edition. 3ª. ed. Cambridge, Massachusetts: The MIT Press, 2008. cap. 1, p. 13-32. ISBN 9780262083645.

SISMONDO, S. **An introduction to science and technology studies**. 2ª. ed. Oxford: Wiley-Blackwell, 2010. 256p. ISBN 9781405187657.

TAULLI, T. **Artificial Intelligence Basics: A Non-Technical Introduction**. 1ª. ed. Berkeley, CA: Apress, 2019. XII, 187 p. ISBN 9781484250273.

TURING, A. M. Computing Machinery and Intelligence. **Mind**: Oxford University Press on behalf of the Mind Association, Oxford, v. LIX, n. 236, p. 433-460, 1 out. 1950. DOI 10.1093/mind/LIX.236.433. Disponível em: <https://academic.oup.com/mind/article/LIX/236/433/986238>. Acesso em: 8 abr. 2023.

WACH, K. *et al.* The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. **Entrepreneurial Business and Economics Review**, [s. l.], ano 2023, v. 11, ed. 2, p. 7-30, 30 jun. 2023. DOI 10.15678/EBER.2023.110201. Disponível em: <https://eber.uek.krakow.pl/index.php/eber/article/view/2113>. Acesso em: 20 set. 2023.

WEST, D. M. **The Future of Work: Robots, Ai, and Automation**. 1^a. ed. Washington, D.C.: Brookings Institution Press, 2018. 219 p. ISBN 9780815732938.

WOLFRAM, S. **What Is ChatGPT Doing ... and Why Does It Work?**. 1^a. ed. Champaign, IL: Wolfram Media Inc, 2023. 102 p. ISBN 9781579550813.

ZHOU, Z. *et al.* Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. **Proceedings of the IEEE**, [s. l.], v. 107, n. 8, p. 1738-1762, Ago. 2019. DOI 10.1109/JPROC.2019.2918951. Disponível em: <https://ieeexplore.ieee.org/document/8736011>. Acesso em: 17 set. 2023.

ZUBOFF, S. **A Era do Capitalismo de Vigilância: A Luta por um Futuro Humano na Nova Fronteira do Poder**. Tradução: George Schlesinger. 1^a. ed. Rio de Janeiro: Editora Intrínseca, 2019. 800 p. ISBN 9786555601442.