

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS HUMANAS
BACHARELADO INTERDISCIPLINAR EM CIÊNCIAS HUMANAS

Bruno Versieux Romano

**INTELIGÊNCIA ARTIFICIAL: EVOLUÇÃO, IMPACTOS E A IMPORTÂNCIA DAS CIÊNCIAS
HUMANAS PARA UM DESENVOLVIMENTO ÉTICO E RESPONSÁVEL**

Artigo apresentado ao Bacharelado Interdisciplinar em Ciências Humanas, da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Bacharel (Trabalho de Conclusão de Curso).
Orientador: Joacir Teixeira de Melo

Juiz de Fora
2024

DECLARAÇÃO DE AUTORIA PRÓPRIA E AUTORIZAÇÃO DE PUBLICAÇÃO

Eu, **BRUNO VERSIEUX ROMANO**, acadêmico do Curso de Graduação Bacharelado Interdisciplinar em Ciências Humanas, da Universidade Federal de Juiz de Fora, regularmente matriculado sob o número **202172012A**, declaro que sou autor do Trabalho de Conclusão de Curso intitulado **INTELIGÊNCIA ARTIFICIAL: EVOLUÇÃO, IMPACTOS E A IMPORTÂNCIA DAS CIÊNCIAS HUMANAS PARA UM DESENVOLVIMENTO ÉTICO E RESPONSÁVEL**, desenvolvido durante o período de **03 DE ABRIL DE 2024** a **03 DE OUTUBRO DE 2024** sob a orientação de **JOACIR TEIXEIRA DE MELO**, ora entregue à UNIVERSIDADE FEDERAL DE JUIZ DE FORA (UFJF) como requisito parcial a obtenção do grau de Bacharel, e que o mesmo foi por mim elaborado e integralmente redigido, não tendo sido copiado ou extraído, seja parcial ou integralmente, de forma ilícita de nenhuma fonte além daquelas públicas consultadas e corretamente referenciadas ao longo do trabalho ou daquelas cujos dados resultaram de investigações empíricas por mim realizadas para fins de produção deste trabalho.

Assim, firmo a presente declaração, demonstrando minha plena consciência dos seus efeitos civis, penais e administrativos, e assumindo total responsabilidade caso se configure o crime de plágio ou violação aos direitos autorais.

Desta forma, na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Juiz de Fora a publicar, durante tempo indeterminado, o texto integral da obra acima citada, para fins de leitura, impressão e/ou download, a título de divulgação do curso de Bacharelado Interdisciplinar em Ciências Humanas e ou da produção científica brasileira, a partir desta data.

Por ser verdade, firmo a presente.

Juiz de Fora, ____ de _____ de _____.

BRUNO VERSIEUX ROMANO

INTELIGÊNCIA ARTIFICIAL: EVOLUÇÃO, IMPACTOS E A IMPORTÂNCIA DAS CIÊNCIAS HUMANAS PARA UM DESENVOLVIMENTO ÉTICO E RESPONSÁVEL

Bruno Versieux Romano¹

RESUMO

A inteligência artificial (IA) tornou-se uma tecnologia revolucionária nos tempos modernos, influenciando vários setores e moldando comportamentos no trabalho, nas relações sociais e cultura. Este estudo examina a evolução histórica da IA, seus principais conceitos, implicações tecnológicas na relação com as ciências humanas e os desafios éticos. A história da IA é analisada, desde uma revisão breve de suas bases teóricas até as aplicações práticas atuais, destacando momentos importantes e avanços significativos. Além disso, a interação entre tecnologia e ciências humanas é discutida, enfatizando a necessidade de uma abordagem interdisciplinar para criar sistemas de IA inclusivos, justos e culturalmente sensíveis. Finalmente, são abordados os impactos da IA no mercado de trabalho, suas implicações para a democracia e a importância de um desenvolvimento que seja ético e que promova a equidade e a justiça social.

PALAVRAS-CHAVE: Inteligência Artificial, História da IA, Ética na IA, Ciências Humanas, Aprendizado de Máquina.

1. INTRODUÇÃO

A inteligência artificial (IA) emergiu como uma das tecnologias mais impactantes da era moderna, influenciando profundamente diversos setores e moldando comportamentos no trabalho, nas relações sociais e cultura. Desde suas origens teóricas até as aplicações práticas contemporâneas, a IA tem demonstrado um potencial transformador que promete redefinir a maneira como vivemos, trabalhamos e interagimos com o mundo. Este trabalho busca explorar a evolução histórica da IA destacando seus principais conceitos, desafios éticos e a importância das ciências humanas no desenvolvimento e aplicação responsável dessas tecnologias.

A trajetória da IA é complexa e multifacetada, começando com mitos antigos sobre autômatos e avançando por marcos significativos, como os trabalhos de Alan Turing e a Conferência de Dartmouth. Com o avanço da tecnologia, a IA se diversificou em áreas como aprendizado de máquina, redes neurais artificiais e processamento de linguagem natural, cada uma contribuindo para inovações que permeiam nossa vida cotidiana.

Contudo, o desenvolvimento e a implementação da IA trazem à tona questões éticas complexas, como viés algorítmico, privacidade e impacto social, que exigem uma reflexão profunda e uma abordagem interdisciplinar. Profissionais das ciências humanas desempenham um papel crucial na navegação desses desafios, fornecendo percepções que ajudam a moldar uma IA que seja tecnicamente avançada, socialmente responsável e culturalmente sensível.

Neste contexto, este trabalho examina a interação entre tecnologia e ciências humanas, demonstrando como a colaboração interdisciplinar pode levar a inovações mais inclusivas e justas. Exploramos também os impactos da IA no mercado de trabalho, as implicações para a democracia e a importância de um desenvolvimento que seja ético e que promova a equidade e a justiça social.

2. HISTÓRICO

A Inteligência Artificial (IA) tem se consolidado como um dos campos mais transformadores da ciência e da tecnologia moderna. Desde suas bases teóricas até suas aplicações práticas em diversas indústrias, a IA demonstra um potencial significativo para alterar profundamente a forma como vivemos e trabalhamos. Esta seção oferece uma visão geral do desenvolvimento histórico da IA e define os conceitos fundamentais que sustentam essa área de estudo.

¹ Graduando em Ciências Humanas pela Universidade Federal de Juiz de Fora – UFJF. romano.bruno@estudante.ufjf.br. Artigo apresentado ao Bacharelado Interdisciplinar em Ciências Humanas como requisito parcial para obtenção do grau de Bacharel. Orientador: Joacir Teixeira de Melo.

A história da IA tem raízes antigas, refletidas em mitos e lendas sobre autômatos. Contudo, o estudo formal da IA começou a ganhar corpo no século XX. Alan Turing, um matemático britânico de grande relevância, é frequentemente citado como um dos pioneiros da computação e da exploração de comportamentos inteligentes em máquinas. Turing é famoso por seu papel na decodificação da máquina de criptografia nazista, Enigma, mas também por suas contribuições teóricas à ciência da computação. Em 1950, ele publicou um artigo fundamental, "Computing Machinery and Intelligence", no qual propôs o famoso "Teste de Turing". Este teste visava avaliar se uma máquina poderia exibir comportamento inteligente de forma indistinguível de um ser humano. O teste funciona assim: um interrogador interage via mensagens de texto com dois participantes, um humano e uma máquina, sem saber qual é qual. A máquina será considerada inteligente se o interrogador não conseguir, de maneira consistente, distinguir quem é a máquina e quem é o humano com base nas respostas fornecidas.

O termo "Inteligência Artificial" foi cunhado em 1956, durante a Conferência de Dartmouth, organizada por John McCarthy, que é amplamente reconhecido como o "pai da IA". Além de cunhar o termo, McCarthy posteriormente desenvolveu a linguagem de programação LISP, amplamente utilizada em programas de IA. Outros participantes notáveis desse evento foram Marvin Minsky, fundador do Laboratório de Inteligência Artificial do MIT e contribuidor no desenvolvimento de redes neurais e teorias cognitivas; Nathaniel Rochester, criador da primeira linguagem de programação da IBM, que tinha interesse em usar computadores para resolver problemas complexos; e Claude Shannon, o "pai da teoria da informação", que trouxe suas ideias sobre comunicação e transmissão de dados para o campo da IA.

Este evento é amplamente considerado o marco oficial da fundação do campo da IA. Os participantes acreditavam que a inteligência humana poderia ser replicada em máquinas, e a conferência visava unir especialistas de diferentes campos para impulsionar a compreensão e a replicação da inteligência em sistemas computacionais. Como explicou McCarthy, a expectativa era que esforços coordenados entre especialistas gerariam avanços significativos na criação de máquinas inteligentes. Segundo McCarthy:

"O projeto de pesquisa de verão sobre inteligência artificial em Dartmouth foi o início da IA como um campo. A proposta era explorar a hipótese de que 'todo aspecto do aprendizado ou qualquer outra característica da inteligência pode, em princípio, ser descrito de maneira tão precisa que uma máquina possa ser feita para simulá-lo.' A conferência reuniu pesquisadores para estudar como programar máquinas para usar a linguagem, formar abstrações e conceitos, resolver tipos de problemas atualmente reservados aos humanos e melhorar a si mesmas." McCarthy (2006)

Nos anos 60 e 70, a IA alcançou alguns dos primeiros sucessos notáveis com a criação de programas capazes de resolver problemas matemáticos e jogar xadrez. No entanto, o campo também enfrentou desafios significativos devido às limitações computacionais da época e à complexidade dos problemas. Esse período de desilusão e redução de financiamento ficou conhecido como o "Inverno da IA" (Crevier, 1993).

Nos anos 80, houve um renascimento do interesse pela IA, impulsionado pelo desenvolvimento de sistemas especialistas, programas projetados para resolver problemas em domínios específicos utilizando regras baseadas no conhecimento humano. Com o avanço dos computadores e a chegada da internet nos anos 90, a IA começou a explorar novos métodos, como aprendizado de máquina (machine learning) e redes neurais artificiais. Russell e Norvig (2010) destacam que esses avanços foram cruciais para revitalizar o campo e abrir novas possibilidades.

O século XXI trouxe avanços significativos, especialmente no campo do aprendizado profundo (deep learning). As melhorias na capacidade de processamento, a disponibilidade de grandes volumes de dados (big data) e o desenvolvimento de algoritmos mais sofisticados permitiram a criação de sistemas de IA com desempenho sem precedentes em tarefas como reconhecimento de fala, visão computacional e processamento de linguagem natural. LeCun, Bengio e Hinton (2015) afirmam que "o aprendizado profundo revolucionou a IA, permitindo avanços em áreas que antes eram consideradas inatingíveis".

3. CONCEITOS RELEVANTES

Nesta seção, são apresentados alguns conceitos-chave de forma introdutória, com o objetivo de facilitar a compreensão dos termos que podem surgir ao longo do trabalho. A abordagem inclui definições essenciais relacionadas à Inteligência Artificial (IA), como algoritmos, aprendizado de máquina e processamento de linguagem natural, entre outros. Embora a explicação seja breve, ela oferece um panorama suficiente para situar os principais elementos do estudo.

3.1. Algoritmo

Um algoritmo é uma sequência finita de instruções claramente definidas, projetadas para resolver um problema ou realizar uma tarefa. Em IA, algoritmos são a base para processar dados e tomar decisões. Por exemplo, um algoritmo de busca pode ser utilizado para encontrar o caminho mais curto entre duas cidades em um mapa. Cormen et al. (2009) explicam: “Um algoritmo é como uma receita em um livro de culinária. Ele especifica uma sequência precisa de instruções que, se seguidas corretamente, resultam na preparação de um prato desejado. Em computação, um algoritmo resolve um problema ao processar uma entrada e gerar uma saída”.

Portanto, em IA, algoritmos são ferramentas fundamentais para que sistemas computacionais resolvam problemas e tomem decisões. Uma Máquina de Turing ou um algoritmo de Markov, por exemplo, são modelos formais que podem ser considerados para a construção de IA, mostrando que a estrutura algorítmica é central para o funcionamento da inteligência artificial.

3.2. Aprendizado de Máquina (Machine Learning)

O aprendizado de máquina é uma subárea da IA que se concentra no desenvolvimento de algoritmos que permitem aos computadores aprender a partir de dados, ajustando-se automaticamente com base nos padrões observados. Existem três tipos principais de aprendizado de máquina:

3.2.1. Aprendizado Supervisionado

No aprendizado supervisionado, o algoritmo é treinado com um conjunto de dados rotulados, ou seja, os dados de entrada são fornecidos juntamente com suas respostas corretas. Um exemplo seria treinar um modelo para reconhecer imagens de gatos, onde cada imagem é previamente classificada como "gato" ou "não gato". O objetivo do aprendizado supervisionado é criar um modelo que generalize a partir desses exemplos, sendo capaz de prever com precisão a saída correta para novos dados. Bishop (2006) afirma: “No aprendizado supervisionado, temos um conjunto de exemplos de treinamento, cada um composto por uma entrada e a saída desejada. O objetivo é que o modelo aprenda a mapear novas entradas para saídas corretas, mesmo que essas entradas não tenham sido vistas anteriormente”. Vale notar que o conjunto de dados de treinamento deve ser diferente do conjunto usado para avaliação, a fim de evitar que o modelo apenas memorize as respostas e não aprenda a generalizar adequadamente.

3.2.2. Aprendizado Não Supervisionado

No aprendizado não supervisionado, o algoritmo recebe dados sem rótulos ou respostas corretas associadas, e seu objetivo é identificar padrões ou estruturas ocultas. Um exemplo seria a segmentação de clientes com base em comportamentos de compra semelhantes, sem que os grupos sejam previamente definidos. Murphy (2012) descreve: “O aprendizado não supervisionado busca descobrir padrões interessantes nos dados, sem a necessidade de um rótulo ou supervisão explícita. Ele tenta encontrar agrupamentos naturais ou relações nos dados sem saber previamente o que está sendo procurado”.

3.2.3. Aprendizado por Reforço

No aprendizado por reforço, o algoritmo aprende a tomar decisões sequenciais interagindo com o ambiente e recebendo feedback na forma de recompensas ou punições. O objetivo é maximizar uma

recompensa cumulativa ao longo do tempo. Um exemplo disso seria um agente de IA aprendendo a jogar xadrez, onde ele ajusta suas ações com base nos resultados obtidos em várias partidas. Sutton e Barto (2018) explicam: "No aprendizado por reforço, um agente interage com seu ambiente, tomando ações e recebendo feedback na forma de recompensas. O objetivo é aprender uma política de ações que maximize a recompensa acumulada ao longo do tempo".

3.3. Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) é um subcampo da IA que se dedica à interação entre computadores e seres humanos por meio da linguagem natural. O PLN permite que máquinas compreendam, interpretem e respondam a linguagem humana de maneira eficaz. Técnicas de PLN são utilizadas em tarefas como tradução automática, análise de sentimentos e sistemas de perguntas e respostas. Por exemplo, o Google Tradutor usa PLN para traduzir textos entre diferentes idiomas. Jurafsky e Martin (2019) afirmam: "O processamento de linguagem natural é essencial para permitir que computadores compreendam e interajam com a linguagem humana. Ele abrange uma ampla gama de técnicas, desde a análise de sintaxe e semântica até a modelagem de discurso e diálogo".

3.4. Chatbots

Os chatbots são programas de computador projetados para simular interações com humanos por meio de conversas em linguagem natural. Utilizando técnicas de processamento de linguagem natural e aprendizado de máquina, os chatbots podem realizar tarefas como atendimento ao cliente, transações financeiras e até fornecer assistência médica. Um exemplo comum é o chatbot de um banco, que responde a perguntas sobre saldos e pode executar transações como transferências. Shawar e Atwell (2007) destacam: "Os chatbots têm o potencial de transformar a maneira como interagimos com a tecnologia, oferecendo uma comunicação mais natural e eficiente. Eles podem automatizar uma variedade de tarefas, desde o suporte ao cliente até serviços de saúde".

4. DESAFIOS ÉTICOS NA ERA DA INTELIGÊNCIA ARTIFICIAL

O desenvolvimento das inteligências artificiais (IAs) está intrinsecamente vinculado ao trabalho dos programadores, que, embora sejam responsáveis diretos pelo código e pela funcionalidade das IAs, não atuam em um contexto de intenções neutras. Como observa André Lemos, "os programadores incorporam suas próprias visões de mundo e interesses nos sistemas que desenvolvem" (Lemos, 2013). Isso significa que as escolhas técnicas e éticas feitas durante o desenvolvimento de algoritmos podem refletir vieses pessoais e objetivos particulares, que nem sempre estão alinhados com os interesses da sociedade em geral. Esse contexto cria um ambiente onde as decisões de programação podem ser guiadas tanto por interesses individuais quanto pelas pressões e diretrizes das instituições para as quais trabalham.

Além disso, a dominação do desenvolvimento de IAs por grandes empresas levanta preocupações significativas sobre a concentração de poder e a priorização do lucro sobre o bem comum. Conforme destaca Sérgio Amadeu da Silveira, "a lógica capitalista das grandes corporações de tecnologia direciona os esforços de inovação para maximizar os lucros, muitas vezes em detrimento de princípios éticos e sociais" (Silveira, 2017). Isso implica que as empresas, ao focarem em retornos financeiros, podem desconsiderar os impactos sociais e éticos das tecnologias que desenvolvem. Portanto, a combinação de objetivos individuais dos programadores e o imperativo lucrativo das empresas cria um cenário onde o desenvolvimento de IAs pode estar desalinhado com as necessidades e os valores da sociedade em geral, exigindo uma governança mais rigorosa e uma reflexão ética contínua sobre o papel dessas tecnologias na nossa vida cotidiana.

A inteligência artificial (IA) estabeleceu-se como uma tecnologia crucial, moldando significativamente diversas esferas da sociedade moderna e provando sua utilidade em uma ampla gama de aplicações. Um exemplo marcante dessa integração é o site Hugging Face, que abriga uma vasta coleção de mais de 400.000 modelos de IA, democratizando o acesso a tecnologias avançadas e facilitando a implementação de soluções inteligentes. No Hugging Face, encontramos principais tipos de IA, como modelos de processamento de linguagem natural (NLP) que são empregados em tradução automática, análise de sentimentos e geração de texto; modelos de visão computacional que são usados para reconhecimento de imagem, detecção de objetos

e análise de vídeo; e modelos de aprendizado por reforço, que têm aplicação em áreas como robótica e jogos. Esta diversidade de modelos e suas aplicações práticas ilustra como a IA não apenas veio para ficar, mas está continuamente evoluindo e expandindo suas capacidades para atender às demandas de um mundo cada vez mais orientado pela tecnologia, consolidando-se como uma ferramenta indispensável para o futuro.

O ChatGPT, como uma manifestação da inteligência artificial, exemplifica a versatilidade e a expansão contínua dessa tecnologia. Baseado na arquitetura GPT (Generative Pre-trained Transformer), o ChatGPT é um modelo de linguagem treinado em uma vasta quantidade de texto, o que lhe permite entender e gerar respostas humanas naturalistas em uma variedade de tópicos. Ao interagir com usuários, o ChatGPT analisa o contexto e as informações fornecidas para oferecer respostas relevantes e úteis. Esta capacidade de compreensão e produção de linguagem natural torna o ChatGPT uma ferramenta valiosa em diversas aplicações, desde assistência ao cliente até geração de conteúdo criativo. Essas duas iniciativas usadas como exemplo, tais como outras não citadas, são tidas por entusiastas da tecnologia, como “democratizadoras da inteligência”, que colocam uma inteligência que está se tornando maior que a nossa, no sentido de uma compilação sem precedentes de informações, nas mãos das pessoas. Mas existem desafios éticos que precisam ser levados em consideração.

O enviesamento político no desenvolvimento de inteligências artificiais (IAs) representa uma preocupação significativa devido ao potencial de perpetuar e ampliar desigualdades sociais e políticas. Esse problema surge quando os algoritmos são influenciados por preconceitos, opiniões ou interesses políticos dos desenvolvedores, resultando em decisões discriminatórias ou injustas. Um exemplo notório desse enviesamento é observado em sistemas de reconhecimento facial que tendem a ter taxas de erro mais altas para pessoas de certos grupos étnicos, devido à falta de diversidade nos conjuntos de dados utilizados para treiná-los. Aplicativos como o Faception ilustram vividamente os perigos do uso de leitura biométrica para determinar características pessoais complexas. O Faception alega poder identificar traços de personalidade, orientação sexual e propensão para o crime com base em análise facial. Por exemplo, ele afirmou ser capaz de detectar “terroristas” ou “criminosos” com base em características faciais específicas. No entanto, essas alegações foram amplamente criticadas por serem imprecisas, arbitrárias e altamente estigmatizantes. O uso desse tipo de aplicativo não apenas reflete um retorno aos tempos da frenologia, onde características físicas eram erroneamente associadas a traços de personalidade, mas também pode ter sérias ramificações sociais, reforçando estereótipos prejudiciais e contribuindo para a marginalização de grupos minoritários.

Portanto, aplicativos que tentam determinar a orientação sexual, criminalidade ou outras características com base em leitura biométrica são perigosos, pois perpetuam estereótipos e discriminação, similarmente à pseudociência da frenologia do século XIX. Esses aplicativos simplificam a complexidade da identidade humana e promovem conclusões errôneas e injustas, alimentando preconceitos e estigmatização. Além disso, podem ser usados para discriminar grupos minoritários, aumentando disparidades sociais e violando direitos humanos.

Além disso, a IA tem sido utilizada para manipular a percepção pública, disseminar desinformação ou influenciar eleições, apresentando conteúdos personalizados que reforçam crenças políticas específicas. Esse fenômeno, conhecido como bolha de filtro, cria uma polarização política e uma fragmentação da sociedade, minando os princípios democráticos de debate aberto e pluralidade de opiniões.

Um outro problema que já estamos acompanhando, é a proliferação das fake news, impulsionada pelas tecnologias de IA. As notícias falsas têm o potencial de desestabilizar a sociedade ao disseminar desinformação e manipular a opinião pública. A IA facilita a criação de conteúdos falsos extremamente convincentes, incluindo textos, imagens e vídeos, tornando difícil distinguir o real do fabricado. André Lemos destaca que “a proliferação de conteúdos falsos e manipulados, impulsionados por algoritmos de IA, coloca em risco a integridade da informação e ameaça a confiança pública nas instituições e nos meios de comunicação” (Lemos, 2020). Essa desinformação não apenas ameaça a democracia, mas também compromete a coesão social ao alimentar teorias da conspiração e polarizar o debate público. Para combater esse problema, Lisandra Vieira argumenta que o combate eficaz à desinformação deve envolver tanto avanços tecnológicos quanto uma transformação educacional e cultural que torne os cidadãos mais conscientes, críticos e preparados para lidar com o ambiente informacional contemporâneo:

“A luta contra as fake news exige não apenas tecnologias que detectem e neutralizem informações falsas, mas também uma educação crítica que capacite os cidadãos a reconhecerem e

questionarem a veracidade das informações que consomem. Isso envolve a promoção de habilidades de pensamento crítico, a capacidade de diferenciar fontes confiáveis de não confiáveis, e a conscientização sobre o impacto social e político da desinformação. Somente através de um esforço conjunto que inclui a inovação tecnológica e a educação cívica podemos mitigar os danos causados pela disseminação de notícias falsas e preservar a integridade do espaço público e do debate democrático" (Vieira, 2019)

Vieira destaca que o combate às fake news requer não apenas soluções tecnológicas, mas também uma transformação educacional que inclua esses aspectos das ciências humanas. A integração desses conhecimentos permite aos cidadãos desenvolver um senso crítico mais aguçado e uma maior sensibilidade para detectar manipulações, desinformações e distorções da realidade. Portanto, o fortalecimento das ciências humanas é fundamental para criar uma sociedade mais consciente e resistente aos impactos negativos da desinformação.

Portanto, é crucial rejeitar e condenar tais práticas, enfatizando a importância da ética e da responsabilidade no desenvolvimento e uso da inteligência artificial. Regulamentações rigorosas são necessárias para proteger os direitos individuais e coletivos, garantindo que a IA seja usada de maneira justa, transparente e respeitosa da diversidade humana.

5. INTEGRAÇÃO INTERDISCIPLINAR

A atuação dos profissionais das ciências humanas no desenvolvimento de inteligências artificiais (IA) envolve diversas questões fundamentais, como a colaboração interdisciplinar, a imparcialidade, a ética, a sensibilidade cultural e o foco no usuário. A colaboração entre desenvolvedores de IA e especialistas em ciências humanas é crucial para criar sistemas éticos, inclusivos e responsáveis. Almeida (2019) destaca que "a colaboração interdisciplinar entre ciências humanas e tecnológicas pode levar a inovações responsáveis em IA." Silveira (2017) reforça que essa colaboração é essencial para construir IA mais inclusivas e justas, com especialistas em diversas áreas contribuindo para soluções éticas e sensíveis aos desafios humanos.

O viés nos algoritmos de IA é um desafio significativo que pode levar a resultados injustos, afetando desproporcionalmente certos grupos. Silveira (2017) enfatiza que "o viés na IA é um desafio significativo que pode ser mitigado por meio da colaboração com ciências sociais." Pasquale (2015) ressalta a importância da transparência em sistemas de IA para promover justiça e responsabilidade, enquanto Silva (2021) destaca que "a busca pela justiça na IA pode ser alcançada através de uma abordagem multidisciplinar, envolvendo ciências humanas."

Considerações éticas e morais são centrais para o desenvolvimento responsável de IA. Sousa (2021) afirma que "a ética aplicada à IA é essencial para garantir que as tecnologias sejam desenvolvidas com responsabilidade e justiça." Silveira (2017) destacam a importância de incorporar princípios éticos e de uma abordagem interdisciplinar para lidar com questões de responsabilidade moral na IA. Floridi & Cowls (2021) reforçam que "a ética da IA é um campo que se baseia em filosofia moral, sociologia e outras ciências humanas para lidar com questões complexas relacionadas à justiça, responsabilidade e transparência."

A sensibilidade cultural é também crítica no desenvolvimento de IA. Canclini (2013) afirma que "a diversidade cultural deve ser considerada no desenvolvimento da IA para garantir a inclusão de diferentes perspectivas." Machado (2019) enfatiza a importância de incorporar ética cultural na IA para evitar que a tecnologia ofenda ou exclua determinados grupos culturais.

Por fim, projetar IA com uma abordagem centrada no usuário é essencial para criar sistemas intuitivos e eficazes. Lemos (2013) destaca que "o design centrado no usuário na IA requer uma compreensão profunda do comportamento e das necessidades humanas." Capurro (2013) ressalta a importância da interação humano-computador e do design de experiências humanas para criar sistemas de IA eficazes e intuitivos, respeitando questões éticas e de privacidade.

6. A CONTRIBUIÇÃO DAS CIÊNCIAS HUMANAS NO AVANÇO DAS INTELIGÊNCIAS ARTIFICIAIS

As ciências humanas desempenham um papel fundamental no desenvolvimento das IAs. É somente através de um trabalho interdisciplinar de parceria entre engenheiros de software, escritores de códigos, designers e demais desenvolvedores com os profissionais provenientes das ciências humanas que podemos alcançar inteligências artificiais que se pareçam mais conosco, ou pelo menos, que não sejam nocivas à humanidade. De acordo com Barfield e Pagallo (2018), "insights de disciplinas como sociologia, psicologia e antropologia podem informar o design da IA para criar sistemas que estejam mais alinhados com o comportamento e expectativas humanos." De forma resumida, as ciências humanas, abrangendo disciplinas como psicologia, sociologia, antropologia e filosofia, estudam as complexidades do comportamento humano, das sociedades e das culturas. Esses campos são cruciais para o desenvolvimento da inteligência artificial (IA) porque fornecem insights sobre a cognição humana, as dinâmicas sociais, as considerações éticas e os contextos culturais, todos essenciais para criar sistemas de IA que sejam eficazes, éticos e benéficos para a sociedade. A seguir, alguns insights aprofundando um pouco nas questões concernentes a cada um desses campos das ciências humanas citados anteriormente, sobre como eles são relevantes no desenvolvimento das IAs.

6.1. Sociologia

A sociologia examina como os indivíduos interagem dentro de grupos e a estrutura das sociedades. Essa disciplina ajuda os desenvolvedores de IA a entender as normas sociais e os comportamentos coletivos, dinâmicas sociais além de enxergar os desafios enfrentados pelos diversos grupos sociais, o que é de vital importância para criar sistemas de IA que possam operar em contextos sociais. Por exemplo, pesquisas sociológicas sobre padrões de uso de redes sociais informaram os algoritmos que alimentam os sistemas de recomendação. Segundo Gillespie (2014), "Os algoritmos são agora componentes chave de nossa infraestrutura digital, com impactos significativos na vida social". Entender os impactos sociais da IA, como a substituição de empregos pela automação, exige insights sociológicos para abordar e mitigar as consequências negativas. Heloisa Almeida observa que os impactos sociais da IA, particularmente nos mercados de trabalho, demandam uma perspectiva sociológica crítica para garantir que a tecnologia sirva ao bem comum (Almeida, 2019). Já Lemos (2013) enfatiza a importância das ciências sociais na IA, uma vez que elas ajudam a compreender os impactos sociais e culturais da tecnologia. Por fim, o diálogo entre pesquisadores de ciências humanas e desenvolvedores de IA pode levar a soluções mais eficazes e sensíveis aos desafios humanos.

6.2. Psicologia

A psicologia fornece conhecimentos fundamentais sobre cognição, emoção e comportamento humanos. Teorias psicológicas sobre aprendizado e memória têm sido instrumentais no desenvolvimento de algoritmos de aprendizado de máquina. A psicologia cognitiva inspirou modelos de redes neurais que imitam processos de aprendizado humano. Como Pantic e Rothkrantz (2000) destacam, "A interpretação de emoções a partir de expressões faciais pode ser significativamente aprimorada pela compreensão dos princípios psicológicos". Além disso, os insights psicológicos sobre a interação humano-computador (IHC) podem melhorar a usabilidade dos sistemas de IA. Um exemplo é o uso de princípios psicológicos no design de chatbots baseados em IA para tornar as interações mais naturais e intuitivas. A psicóloga brasileira Ana Bock (2018) afirma: "A integração de princípios psicológicos no desenvolvimento de IA ajuda a criar sistemas que podem compreender e responder às emoções humanas de maneira mais eficaz".

6.3. Antropologia

O foco da antropologia na diversidade cultural e nas sociedades humanas fornece à IA uma estrutura para entender como as práticas e contextos culturais específicos e interseccionalidades influenciam o uso da tecnologia. A pesquisa antropológica pode orientar o desenvolvimento de sistemas de IA que respeitem e se adaptem a diferentes normas e práticas culturais. Por exemplo, a tecnologia de reconhecimento facial deve levar em conta características faciais diversas entre as populações, e os insights antropológicos podem ajudar a tornar esses sistemas mais precisos e inclusivos. Boellstorff (2008) enfatiza que "entender os contextos

culturais nos quais as tecnologias são usadas é crucial para desenvolver sistemas de IA inclusivos". A antropóloga brasileira Manuela Carneiro da Cunha (2009) afirma: "Incorporar insights antropológicos no desenvolvimento da IA garante que essas tecnologias sejam culturalmente sensíveis e adaptáveis, promovendo aplicações mais inclusivas e eficazes".

6.4. Filosofia

A filosofia contribui para a IA ao abordar questões éticas fundamentais e a natureza da inteligência e da consciência. Os debates filosóficos sobre a ética da IA, como aqueles concernentes à privacidade, autonomia e justiça, são cruciais para desenvolver diretrizes e regulamentações que governem a implantação da IA. Como Bostrom e Yudkowsky (2014) observam, "As considerações éticas são primordiais no desenvolvimento e implementação de tecnologias de IA para garantir que beneficiem a humanidade". O dilema do bonde, um clássico dilema ético que consiste em uma situação onde um bonde desgovernado está prestes a atropelar cinco pessoas amarradas em uma linha de trem. Você está ao lado de uma alavanca que pode desviar o bonde para outra linha, onde está amarrada apenas uma pessoa. O dilema é: você deve puxar a alavanca, sacrificando uma pessoa para salvar cinco, ou não fazer nada, permitindo que as cinco pessoas morram, é frequentemente usado para explorar as decisões morais que veículos autônomos devem tomar em situações críticas. Além disso, questões da filosofia da mente sobre a natureza da consciência e do pensamento influenciam a pesquisa de IA na criação de máquinas com compreensão semelhante à humana. O filósofo brasileiro Roberto Mangabeira Unger (2016) destaca: "O exame filosófico da ética da IA é essencial para orientar o desenvolvimento responsável da tecnologia, garantindo que ela esteja alinhada com os valores humanos e o bem-estar social". Além disso, questões relacionadas à autonomia, justiça e responsabilidade dos sistemas de IA devem ser discutidas com base em perspectivas filosóficas sólidas. A filosofia pode também orientar a elaboração de códigos de conduta para pesquisadores e desenvolvedores de IA.

7. CONCLUSÕES

O futuro das inteligências artificiais aponta para o desenvolvimento de capacidades cada vez mais sofisticadas, como a teoria da mente e a autoconsciência. A teoria da mente envolve a capacidade de uma IA entender e modelar as crenças, intenções e desejos de outros seres, permitindo interações mais naturais e empáticas. A autoconsciência, por outro lado, representaria uma etapa em que a IA não apenas entenderia o mundo ao seu redor, mas também teria uma compreensão de si mesma como entidade independente.

Esse avanço, porém, levanta questões profundas sobre a ética e a responsabilidade, uma vez que uma IA com autoconsciência poderia desafiar nossas noções de consciência e dignidade humana. Rafael Capurro enfatiza a importância de uma reflexão ética contínua, afirmando que "a tecnologia, por si só, não pode substituir o julgamento ético e a interação humana essencial em profissões que lidam diretamente com o bem-estar humano" (Capurro, 2009, p. 45).

O impacto da inteligência artificial no mercado de trabalho representa um desafio significativo para diversas profissões especializadas, como professores, profissionais de saúde e outros. A automação e o desenvolvimento de tecnologias avançadas estão transformando a natureza do trabalho, criando um cenário onde tarefas tradicionalmente realizadas por humanos podem ser realizadas por máquinas com eficiência e precisão superiores. As tecnologias de automação e inteligência artificial têm o potencial de transformar significativamente diversas profissões, inclusive aquelas tradicionalmente ocupadas por humanistas. Como aponta Ruy Braga, "as mudanças tecnológicas podem levar a uma reconfiguração do trabalho, exigindo que os profissionais adquiram novas habilidades que complementem a tecnologia, em vez de serem substituídos por ela" (Braga, 2017, p. 145).

No campo da educação, por exemplo, o uso crescente de plataformas de aprendizado online e tutores virtuais baseados em IA coloca em risco a demanda por professores em determinadas funções, especialmente em atividades repetitivas e administrativas. Da mesma forma, na área da saúde, a introdução de sistemas de diagnóstico por IA e robôs cirúrgicos avançados ameaça reduzir a necessidade de certos especialistas médicos, como radiologistas e cirurgiões, cujas tarefas podem ser parcialmente ou totalmente automatizadas. No entanto, como alerta Ruy Braga, "a crescente substituição da força de trabalho humana por máquinas

inteligentes pode resultar em um deslocamento significativo de empregos, aumentando o desemprego e a precariedade laboral, mesmo entre os profissionais altamente qualificados" (Braga, 2017, p. 150).

Além disso, essa transformação tecnológica exige que os trabalhadores adquiram novas competências e adaptem suas habilidades para complementar, em vez de competir com, a tecnologia. A capacidade de aprendizado contínuo e a adaptação a novas funções que envolvam a colaboração com sistemas de IA serão essenciais para mitigar os riscos de desemprego e subemprego, garantindo que os profissionais possam prosperar em um mercado de trabalho em rápida evolução. A questão é particularmente preocupante em áreas que requerem um elevado grau de empatia e julgamento humano, onde a substituição por IA pode levar não apenas a uma perda de emprego, mas também a uma diminuição na qualidade dos serviços prestados, como sugere Capurro, que afirma que "a tecnologia, por si só, não pode substituir o julgamento ético e a interação humana essencial em profissões que lidam diretamente com o bem-estar humano" (Capurro, 2009, p. 45). Portanto, é crucial que haja um esforço coordenado entre governos, empresas e instituições educacionais para garantir uma transição justa e equitativa, promovendo a requalificação e a atualização profissional contínua para enfrentar os desafios impostos pela evolução tecnológica.

Os avanços em inteligência artificial (IA) estão transformando rapidamente diversos setores da sociedade, trazendo consigo uma série de desafios e oportunidades que afetam tanto o mercado de trabalho quanto a qualidade da informação disponível e a evolução futura dessas tecnologias. No mercado de trabalho, a automação e a IA estão substituindo funções que antes eram realizadas por humanos, especialmente em áreas como educação e saúde. Professores, por exemplo, enfrentam a concorrência de plataformas de aprendizado automatizadas e tutores virtuais, enquanto profissionais de saúde lidam com o advento de sistemas de diagnóstico e cirurgia assistida por IA.

Essa transformação pode levar a uma redução na demanda por determinados especialistas, resultando em desemprego e subemprego, mesmo entre aqueles com alta qualificação. Ruy Braga alerta que "a crescente substituição da força de trabalho humana por máquinas inteligentes pode resultar em um deslocamento significativo de empregos, aumentando o desemprego e a precariedade laboral, mesmo entre os profissionais altamente qualificados" (Braga, 2017, p. 150). Além disso, existe um grupo de pessoas que se torna cada vez mais não-empregável, incapaz de acompanhar a rapidez das mudanças tecnológicas e de se adaptar a um mercado que exige habilidades constantemente renovadas e complexas.

A IA também pode exacerbar desigualdades sociais e econômicas já existentes. Maria da Glória Gohn discute que "há um risco significativo de que as vantagens proporcionadas pelas novas tecnologias sejam distribuídas de maneira desigual, beneficiando principalmente aqueles com acesso a recursos tecnológicos avançados e marginalizando ainda mais os grupos vulneráveis" (Gohn, 2011, p. 79). Os profissionais das ciências humanas têm o desafio de desenvolver estratégias que promovam uma distribuição mais justa dos benefícios da IA, garantindo que todos possam usufruir das vantagens dessa revolução tecnológica.

Outro desafio importante é a preservação da identidade cultural em um contexto global cada vez mais influenciado pela IA. A homogeneização cultural é um risco real, e Néstor García Canclini ressalta que "há uma necessidade urgente de manter e promover a diversidade cultural e as tradições locais frente à crescente influência da tecnologia globalizante" (Canclini, 2003, p. 112). Profissionais das ciências humanas devem trabalhar para garantir que a diversidade cultural seja valorizada e preservada, evitando a diluição das identidades locais em um mundo tecnologicamente uniforme.

A educação e a capacitação são áreas cruciais onde os profissionais das ciências humanas enfrentarão desafios significativos. Rosana Guber sugere que "integrar conhecimentos tecnológicos e éticos nos programas educacionais é essencial para formar cidadãos críticos e preparados para os desafios do século XXI" (Guber, 2001, p. 97). É fundamental atualizar os currículos e métodos educacionais para preparar os estudantes para um futuro em que a IA estará presente em todas as esferas da vida.

Além disso, a crescente quantidade de dados gerados pela IA demanda habilidades de interpretação crítica e contextualização. A análise de grandes volumes de dados sem uma perspectiva humanística pode levar a conclusões simplistas ou enviesadas. André Lemos argumenta que os humanistas, com sua capacidade de análise crítica e contextual, serão essenciais para interpretar esses dados de forma que reflitam uma compreensão mais profunda e ética dos fenômenos sociais e culturais (Lemos, 2013).

A participação na tomada de decisões é outra área em que a presença de profissionais das ciências humanas será fundamental. As IAs estão cada vez mais envolvidas em decisões que afetam a vida das pessoas, desde a saúde até a segurança. Garantir que essas decisões sejam justas e inclusivas requer a participação ativa de humanistas em comitês e grupos de trabalho que desenvolvem e aplicam essas tecnologias. Virgílio Afonso

da Silva observa que “a tecnologia pode ameaçar a privacidade e outros direitos fundamentais, e os profissionais das ciências humanas terão que trabalhar arduamente para desenvolver políticas que protejam esses direitos” (Silva, 2009).

A autenticidade das experiências humanas e a prevenção da dependência excessiva da tecnologia são preocupações crescentes à medida que a IA se integra em diversos aspectos da vida cotidiana. Defensores das humanidades devem promover a importância das relações e experiências humanas autênticas, destacando o valor da interação humana em um mundo cada vez mais digitalizado. Pedro Nunes afirma que “garantir que as tecnologias de IA sejam seguras e confiáveis é essencial, e os humanistas devem colaborar com tecnólogos para desenvolver padrões que assegurem essa confiabilidade” (Nunes, 2021).

Por fim, a IA está transformando a comunicação e as interações sociais. Novas formas de comunicação, como chatbots e tradutores automáticos, estão se tornando comuns, e os profissionais das ciências humanas precisam se adaptar a essas mudanças e compreender como elas impactam a sociedade e a cultura. A segurança e confiabilidade dos sistemas de IA também são preocupações importantes. Como afirma Rafael Capurro, “os desafios que a IA impõe à sociedade contemporânea exigem uma reflexão crítica e profunda para garantir que as tecnologias emergentes contribuam para o bem comum” (Capurro, 2012).

Esses desafios refletem a complexidade e a interconexão entre as humanidades e a tecnologia. Os profissionais das ciências humanas desempenharão um papel vital em moldar um futuro em que a tecnologia sirva ao bem-estar humano e à sociedade como um todo. A reflexão sobre os impactos das tecnologias e a promoção de um uso ético e inclusivo da IA serão fundamentais para garantir que os avanços tecnológicos contribuam para o desenvolvimento sustentável e a equidade social.

Sem a atuação dos profissionais das ciências humanas, o desenvolvimento de tecnologias de IA corre o risco de ser guiado apenas por parâmetros técnicos e econômicos, negligenciando a ética, a justiça social e a diversidade cultural. Os humanistas são essenciais para garantir que a IA considere as implicações sociais e culturais de maneira responsável, promovendo alfabetização digital crítica, elaborando políticas públicas para proteção dos direitos humanos e defendendo a importância das interações humanas autênticas. Sua integração é crucial para moldar um futuro onde as inovações tecnológicas beneficiem a sociedade como um todo, promovendo equidade social e desenvolvimento sustentável. Sem essa intervenção, a IA pode aumentar desigualdades, marginalizar grupos vulneráveis e comprometer a qualidade das interações humanas. Portanto, é necessário um esforço coordenado entre tecnólogos e humanistas para garantir que o progresso tecnológico seja guiado por princípios éticos e humanitários, servindo ao bem-estar coletivo.

REFERÊNCIAS

ALMEIDA, Heloísa. **Sociologia e Inteligência Artificial no Brasil**. São Paulo: EdUSP, 2019.

BARFIELD, W.; PAGALLO, U. **The Cambridge Handbook of Artificial Intelligence**. Cambridge: Cambridge University Press, 2018.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Berlin: Springer, 2006.

BLODGETT, S. L.; O'CONNOR, B. **Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English**. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.

BOCK, A. **Psicologia: Uma Introdução ao Estudo de Psicologia**. Petrópolis: Editora Vozes, 2018.

BOELLSTORFF, T. **Coming of Age in Second Life: An Anthropologist Explores the Virtually Human**. Princeton: Princeton University Press, 2008.

BOSTROM, N.; YUDKOWSKY, E. **The Ethics of Artificial Intelligence**. In: GUENTHNER, F. B.; GABBAY, D. G. (Eds.). *Handbook of the Philosophy of Science*. Amsterdam: Elsevier, 2014.

CANCLINI, Néstor García. **A Globalização Imaginada**. Rio de Janeiro: Zahar, 2003.

- CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. **Introduction to Algorithms**. Cambridge: MIT Press, 2009.
- CREVIER, D. **AI: The Tumultuous History of the Search for Artificial Intelligence**. New York: Basic Books, 1993.
- FLORIDI, L.; COWLS, J. **The Cambridge Handbook of Ethics and AI**. Cambridge: Cambridge University Press, 2021.
- GILLESPIE, T. **The Relevance of Algorithms**. In: GILLESPIE, T.; BOCZKOWSKI, P. J.; FOOT, K. A. (Eds.). *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge: MIT Press, 2014.
- GOHN, M. G. **Educação Não Formal e Cultura Política: Impactos na Cidadania e Democracia**. São Paulo: Cortez Editora, 2011.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.
- GUBER, R. **La Etnografía: Método, Campo y Reflexividad**. Buenos Aires: Siglo XXI Editores, 2001.
- HAGENDORFF, T. **The Ethics of AI Ethics: An Evaluation of Guidelines**. *Minds and Machines*, v. 30, p. 99-120, 2020.
- JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**. London: Pearson, 2019.
- LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep Learning**. *Nature*, v. 521, n. 7553, p. 436-444, 2015.
- LEMOS, André. **Cibercultura: tecnologia e vida social na cultura contemporânea**. Porto Alegre: Sulina, 2013.
- LEMOS, André. **Fake news e a manipulação da opinião pública na era digital**. Porto Alegre: Sulina, 2020.
- MACHADO, Irene. **Inteligência artificial e comunicação: uma abordagem crítica**. São Paulo: Editora UNESP, 2019.
- MEHRABI, N.; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. **A Survey on Bias and Fairness in Machine Learning**. *ACM Computing Surveys (CSUR)*, v. 54, n. 6, p. 1-35, 2021.
- MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**. Cambridge: MIT Press, 2012.
- NUNES, Pedro. **Segurança e privacidade na era da inteligência artificial**. São Paulo: Alameda, 2021.
- PANTIC, M.; ROTHKRANTZ, L. J. M. **Automatic Analysis of Facial Expressions: The State of the Art**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 12, p. 1424-1445, 2000.
- PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information**. Cambridge: Harvard University Press, 2015.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Upper Saddle River: Prentice Hall, 2010.
- SHNEIDERMAN, B. **Human-Centered AI**. Oxford: Oxford University Press, 2020.
- SHAWAR, Bayan Abdulrahman; ATWELL, Eric. **Chatbots: Are They Really Useful?** In: *Linguistic Engineering*, v. 3, 2007.

SICILIANO, B.; KHATIB, O. **Springer Handbook of Robotics**. Berlin: Springer, 2016.

SILVA, Virgílio Afonso da. **Direitos Fundamentais: Conteúdo Essencial, Restrições e Eficácia**. São Paulo: Malheiros Editores, 2009.

SZELISKI, R. Computer Vision: **Algorithms and Applications**. Berlin: Springer, 2010.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. Cambridge: MIT Press, 2018.

TURING, A. M. **Computing Machinery and Intelligence**. *Mind*, v. 59, n. 236, p. 433-460, 1950.

UNGER, R. M. **The Critical Legal Studies Movement: Another Time, a Greater Task**. London: Verso Books, 2016.

VIEIRA, Lisandra. **Fake news e o impacto nas eleições: um olhar crítico**. São Paulo: Alameda, 2019.