



Universidade de Brasília

Faculdade de Ciências e Tecnologias em Engenharia

Introdução à Inteligência Artificial

Maciel Ferreira Custódio Júnior

Matrícula: 19/0100087

Disciplina: Inteligência Artificial

Professor: Fabiano Araujo Soares

Brasília
2025

1 Introdução

A inteligência artificial (IA) é definida por Russell e Norvig (2021) como uma área que se baseia no conceito de agentes inteligentes que recebem percepções do ambiente e realizam ações. Na prática, cada agente segue um conjunto de regras que transformam o que foi percebido em decisões sobre como agir. A definição enfatiza a autonomia e a capacidade de tomada de decisão como elementos centrais da inteligência artificial.

Russell e Norvig (2021) também destacam a abrangência da IA, que envolve desde aprendizagem e percepção até tarefas específicas como jogos de xadrez, criação de poesia e diagnóstico de doenças. Embora as aplicações atuais sejam amplas, a trajetória da IA é recente. Iniciada após a Segunda Guerra Mundial, alternou entre períodos de otimismo e frustração, os chamados "invernos da IA". Décadas depois, ganhou novo fôlego com os foundation models, descritos por Huyen (2024) como capazes de realizar tarefas antes consideradas impossíveis.

Os impactos sociais e econômicos gerado pela inteligência artificial também são evidentes. O AI Index Report 2025 registra que a IA está cada vez mais incorporada na vida cotidiana. Algumas evidências disso são o número crescente de dispositivos médicos habilitados por IA aprovados pela FDA, que alcançou 223 dispositivos em 2023 contra apenas 6 em 2015 e também o grande número de viagens realizadas com veículos autônomos atualmente. O documento ressalta, também, o aumento de investimentos no setor que teve um aumento significativo de 2023 para 2024 com destaque para a IA Generativa, que teve um impulso particularmente forte (GIL; PERRAULT, 2025).

A disciplina de Inteligência Artificial apresentou diversas dimensões da IA, desde sua concepção até marcos históricos, como o Teste de Turing, os ciclos de "booms" e "invernos" e a ascensão do deep learning. Também foram discutidas definições, abordagens, benefícios e riscos éticos da IA. Dessa forma, esta resenha tem como objetivo apresentar uma síntese crítica dos fundamentos estudados, considerando seus impactos sociais, éticos e econômicos. Para tanto, serão utilizadas fontes complementares a fim de aprofundar o tema de inteligência artificial.

2 Resumo

Esta resenha apresenta uma análise crítica sobre a inteligência artificial (IA), tomando como base os temas discutidos em aula e ampliando-os com pesquisa bibliográfica em autores como Russell e Norvig (2021), Huyen (2024) e no AI Index Report 2025 (GIL; PERRAULT, 2025). O texto aborda os fundamentos e definições da área, sua evolução histórica, o estado atual e suas principais aplicações. São analisados tópicos relevantes, incluindo riscos, como vieses, privacidade, desinformação e benefícios, como personalização de serviços, eficiência, apoio à decisão e inclusão social. O trabalho também discute os impactos éticos e sociais do desenvolvimento e uso da inteligência artificial, destacando a importância da criação de leis e regulamentações específicas para a área.

3 Análise Crítica

3.1 Panorama Inicial da IA

Russell e Norvig (2021) afirmam que a inteligência artificial abrange uma diversidade de subcampos, se concretizando como um campo universal. Essa definição é bastante forte porque deixa claro que a IA pode estar presente nas mais diversas áreas, mas também apresenta um problema, a dificuldade em definir claramente onde ela começa e termina. Huyen (2024) também corrobora com a ideia da vasta abrangência da IA ao afirmar que os modelos atuais permitem novas aplicações que antes eram consideradas impossíveis.

O IA Index Report 2025 (GIL; PERRAULT, 2025) traz uma relação dessa visão com o cotidiano. Afirmando que a IA está presente cada vez mais no nosso cotidiano. Desde a saúde até o transporte, a inteligência artificial está se tornando um complemento diário em nossas vidas. O que deixa claro que a IA não é mais apenas pesquisa, mas já se tornou uma tecnologia que afeta grande parte da população. Dessa forma, é necessário entender os riscos e benefícios associados ao seu uso.

Segundo Russell e Norvig (2021) a IA possui potencial para libertar a humanidade de trabalhos repetitivos e acelerar pesquisas científicas, podendo auxiliar na descoberta de curas para doenças e soluções para problemas climáticos e de escassez de recursos. No entanto, os autores também alertam para potenciais riscos, como criação de armas letais, falta de transparência e problemas com privacidade de dados. O IA Index Report 2025 confirma essas preocupações, mostrando que empresas estão atentas aos riscos ligados à IA. A pesquisa apontou como principais preocupações os riscos financeiros, relacionados a possíveis perdas econômicas, os de marca e reputação, que afetam diretamente a confiança pública e os riscos associados a privacidade e segurança de dados.

3.2 Definições de Inteligência e de IA

O conceito de inteligência é bastante diverso, Russell e Norvig (2021) afirmam que não existe uma definição única e universal. Esse ponto é importante pois apesar de deixar clara a dificuldade em estabelecer um consenso, também abriu espaço para diferentes abordagens sobre o tema ao longo da história, sendo as quatro principais listadas abaixo:

1. **Agir como humano:** essa abordagem está associada ao teste de turing. Os autores explicam que o teste proposto por Alan Turing na década de 50, foi criado com o objetivo de oferecer uma definição satisfatória de inteligência (RUSSELL; NORVIG, 2021). Apesar de ser um teste prático e observável, a Stanford Encyclopedia of Philosophy (OPPY; DOWE, 2021) afirma que ele se limita a avaliar apenas a imitação do comportamento humano evitando a discussão filosófica do que é inteligência.
2. **Pensar como humano:** visão que busca embasamento na ciência cognitiva e na psicologia reunindo modelos computacionais de IA e técnicas da psicologia para reunir

teorias precisas e testáveis da mente humana (RUSSELL; NORVIG, 2021). Entretanto, há limitações nessa abordagem, vez que ainda sabemos muito pouco sobre como a mente humana realmente funciona.

3. **Pensar racionalmente:** essa linha de pensamento vem do estudo da lógica. Segundo Russell e Norvig (2021), Aristóteles foi um dos primeiros a tentar codificar o "pensamento correto". O uso de silogismos fornecia padrões de argumentos que levavam sempre a conclusões corretas dadas premissas corretas. Apesar de trazer avanços, sua limitação é lidar com incertezas e ambiguidades, já que pensamento humano possui contradições e incertezas, o que distancia as regras de lógica utilizadas nessa abordagem do pensamento real.
4. **Agir racionalmente:** mais aceita atualmente, essa abordagem define a IA como o estudo de agentes que recebem percepções do ambiente e executam ações com base nessas percepções. Dessa forma, a inteligência é definida pela capacidade de agir de forma eficaz em relação a um objetivo. Porém, a limitação é que um agente pode ser "racional" segundo um objetivo mal definido, podendo gerar resultados prejudiciais (RUSSELL; NORVIG, 2021).

3.3 Histórico da IA

Os primeiros vestígios da ideação da inteligência artificial são antigos e têm origens nas narrativas mitológicas. Por exemplo, o mito de Talos (700 a.C) que apresenta um homem gigante de bronze construído por Hefesto para proteger Creta, pode ser considerada uma das primeiras concepções de um robô (Stanford University, 2019). Mais recentemente, no campo da ficção, Frankenstein de Mary Shelley surge como uma narrativa desse debate, levantando discussões sobre ética e responsabilidade na criação de seres artificiais que tomam consciência.

Como estudo científico, a IA começou a se estruturar no século XX, com a proposta do primeiro modelo matemático de redes neurais em 1943 por McCulloch e Pitts. Posteriormente Donald Hebb formulou a regra de aprendizado sináptico em 1949, seguido por Alan Turing em 1950, quando criou o Teste de Turing e escreveu o artigo *Computing Machinery and Intelligence*, considerado uns dos principais fundamentos da IA moderna. O momento decisivo ocorreu em 1956 com a conferência de Dartmouth, onde o próprio termo Artificial Intelligence foi definido por John McCarthy (RUSSELL; NORVIG, 2021).

Nas décadas seguintes, a área passou por períodos de entusiasmo e frustração, conforme recordam Russell e Norvig (2021), a IA alternou entre períodos de otimismo seguidos por decepções e perdas de financiamento seguidos por novas abordagens, sucesso e otimismo renovado. Nas décadas de 70 e 80, os holofotes recaíram sobre os sistemas especialistas, programas que codificavam conhecimento humano em regras.

A partir de 1990, houve o retorno das redes neurais e o crescimento dos métodos probabilísticos, que proporcionaram maior flexibilidade e capacidade de trabalhar com incertezas. O que abriu caminho para a era do Big Data e do Deep Learning. Sobre o fenômeno

do Big Data, Russell e Norvig (2021) afirmam que o grande avanço no poder computacional e a criação da *World Wide Web* facilitaram o surgimento de conjuntos de dados muito grandes, conhecidos como Big Data. A disponibilidade desses dados o surgimento do aprendizado de máquina ajudaram a IA a recuperar a atividade comercial. Sobre o Deep Learning os autores explicam que a competição ImageNet em 2012, que exigia a classificação de imagens em uma entre mil categorias, foi um marco para área. Na competição surgiu um sistema de deep learning criado no grupo de Geoffrey Hinton, na Universidade de Toronto, que desempenhou uma melhora significativa em relação aos sistemas anteriores. Esse avanço estendeu-se para várias áreas da IA, como tarefas de fala, tradução, diagnósticos e jogos.

Esse percurso mostra que os avanços recentes na área da inteligência artificial não foram repentinos, mas sim o resultado de décadas de estudos. Permitindo que a IA evoluísse e se consolidasse até o estágio em que a conhecemos hoje.

3.4 Estado Atual da IA

Nos últimos anos a inteligência artificial foi marcada por avanços técnicos significativos, tornando-a uma grande tendência na atualidade. O AI Index Report 2025 registra que a IA está cada vez mais incorporada no nosso cotidiano, desde a saúde, até nos transportes (GIL; PERRAULT, 2025). Esse movimento de ascensão da IA, segundo Huyen (2024) foi impulsionado pelo surgimento dos modelos fundamentais, que ampliaram a escala e generalidade dos sistemas, permitindo que soluções antes impossíveis de serem criadas agora fossem possíveis.

O AI Index Report 2025 mostra que a IA está presente nos mais diversos setores da sociedade, com impactos diretos e significativos em nossas vidas. Na área da saúde, por exemplo, são citados dispositivos médicos habilitados por IA, cada vez mais presentes na medicina e também dados os sintéticos, que têm potencial para ajudar a identificar determinantes sociais de saúde, melhorar previsões clínicas e descobrir novos fármacos. No setor de transportes, é destacado que os carros autônomos não são mais experimentais e realizam milhares de viagens autônomas por semana. O uso da IA na área da educação também é crescente, cerca de dois terços dos países já oferecem ou planejam oferecer o ensino da ciência da computação na educação básica, número expressivamente maior que em 2019. A IA também evoluiu no setor da segurança, percebida pelo aumento significativo de sistemas de reconhecimento facial pela polícia e por varejistas, mostrando seu potencial para reduzir o crime e melhorar a segurança pública e privada (GIL; PERRAULT, 2025).

Do ponto de vista econômico, o AI Index Report 2025 mostra que os números são expressivos, destacando o investimento privado em IA nos Estados Unidos, que ultrapassou 100 bilhões de dólares e também o investimento global em IA generativa que teve um aumento de 18,7% em relação ao ano de 2023 (GIL; PERRAULT, 2025).

Por fim, os benefícios e riscos da inteligência artificial permanecem centrais na discussão. O AI Index Report 2025 enfatiza que, ao mesmo tempo em que sistemas de IA oferecem diversos ganhos nas áreas como saúde, transporte, robótica e educação, também levantam preocupações relacionadas à privacidade, ao viés algorítmico, à concentração de poder econômico e à vigilância

em massa (GIL; PERRAULT, 2025). Esses fatos mostram que o avanço da tecnologia não pode ser analisado apenas por suas conquistas técnicas, mas também pelas implicações éticas e sociais que carrega, temas esses que serão abordados mais à frente, juntamente com os benefícios e riscos da inteligência artificial.

3.5 Agentes Inteligentes

O entendimento de agentes é de suma importância do contexto da inteligência artificial. Segundo Russell e Norvig (2021), um agente é qualquer coisa que pode ser vista como percebendo seu ambiente por meio de sensores e atuando sobre o ambiente por meio de atuadores. Essa definição nos permite entender agentes de uma forma ampla. Os autores definem ainda a IA como o estudo de agentes que recebem percepções do ambiente e executam ações com base em suas percepções, situando o conceito estudado na disciplina.

A arquitetura de um agente pode ser descrita com base em sua função, implementação, e forma como interage com o ambiente. Conforme Russell e Norvig (2021), a função do agente é uma descrição matemática que relaciona as possíveis sequências de percepções do agente com as ações que serão tomadas em cada caso. O programa do agente, por sua vez, é a implementação dessa função em código, executando no hardware, de forma que o agente possa executar as ações necessárias.

Dessa forma, podemos identificar quatro elementos fundamentais nos atuadores:

1. **Sensores:** captam informações do ambiente
2. **Atuadores:** permitem a intervenção do agente no ambiente
3. **Programa do agente:** implementa as regras de decisão
4. **Ambiente:** contexto ao qual o agente está inserido

Como citado por Russell e Norvig (2021), o comportamento de um agente é estruturado por ciclos de percepção, no qual o agente percebe o ambiente por meio de sensores e age sobre o ambiente por meio de atuadores. Esse processo pode ser mais simples ou mais complexo, dependendo do programa que implementa a função do agente. Ainda segundo os autores, a noção de racionalidade está presente, porém não pode ser confundida com perfeição. Um agente racional é aquele que faz a coisa certa com base nas percepções até o momento, ou seja, um agente racional escolhe a melhor ação possível com base nas informações disponíveis no momento.

Para analisar o desempenho dos agentes, Russell e Norvig (2021) propõe a estrutura PEAS (Performance measure, Environment, Actuators, Sensors). A medida de desempenho, define os critérios que dizem se o agente está indo bem ou mal. O ambiente refere-se ao contexto em que o agente atua, como estradas, trânsito. Os atuadores, são os mecanismos que o agente tem para agir, em um carro autônomo, por exemplo, estão presente: olante, freios, acelerador, etc. E os sensores, por fim, são os responsáveis por receber as informações do mundo, como câmera, radar e gps.

Além disso, os ambientes podem ser classificados por diferentes propriedades, tais como: totalmente vs parcialmente observável, agente único vs. multiagente, determinístico vs. estocástico, episódico vs sequencial, estático vs. dinâmico, discreto vs. contínuo, conhecido vs. desconhecido. Essas classificações ajudam a entender os desafios que os agentes enfrentam de acordo com o ambiente em que está inserido.

Os agentes também são distinguidos por tipos, cada um com diferentes níveis de sofisticação, conforme Russell e Norvig (2021). São eles:

1. **Reflexo simples:** escolhem suas ações com base apenas na percepção atual, seguindo regras condicionais e ignorando o histórico de percepções.
2. **Reflexo baseado em modelo:** superam as limitações dos agentes de reflexo simples ao lidar com ambientes parcialmente observáveis. Para tal, o agente mantém um estado interno do ambiente conforme seu histórico de percepções, o que permite ao agente refletir aspectos não observados do estado atual.
3. **Baseado em objetivos:** ampliam a capacidade de decisão ao considerar não apenas o estado atual do ambiente, mas também estados desejáveis a serem alcançados. Esse tipo de agente é mais flexível, já que seu comportamento pode ser facilmente alterado para ir a um destino diferente.
4. **Baseado em utilidade:** avaliam não apenas se uma ação leva a um objetivo, mas também a qualidade dos diferentes estados que podem resultar. Essa abordagem permite decisões mais flexíveis, avaliando alternativas em vez de tratá-las como escolhas binárias.

Por fim, Russell e Norvig (2021) apresentam o conceito de agente que aprende, que incorpora formas de melhorar o seu desempenho com base na sua experiência. Esse aprendizado está associado à forma como os estados são representados. Os autores descrevem três possibilidades: atômica, fatorial e estruturada. A representação atômica trata cada estado como uma unidade indivisível. Na fatorial, os estados são descritos por variáveis e atributos. Já a estruturada permite modelar objetos e relações entre eles. Esses diferentes tipos abrem espaços para diferentes formas de raciocínio e aprendizado.

3.6 Ética, Riscos e Benefícios da IA

3.6.1 Ética e Responsabilidade

A ética na área da inteligência artificial começa pela noção de responsabilidade do ponto de vista humano. Russell e Norvig (2021) afirmam que "Robôs são ferramentas que criamos para realizar as tarefas que lhes designamos, e se lhes concedermos personalidade, estamos apenas nos recusando a assumir a responsabilidade pelas ações de nossa propriedade"(tradução), destacando que no final, a responsabilidade é sempre humana, e não da tecnologia em si.

Dilemas éticos mostram como a responsabilidade humana é bastante complexa em cenários práticos de IA. Por exemplo, os dilemas éticos em veículos autônomos, que têm sido estudados pelo experimento *Moral Machine* do MIT (*Massachusetts Institute of Technology*), reuniu milhões de respostas sobre como as pessoas tomariam decisões em cenários de acidentes inevitáveis (MIT Media Lab, 2016). Isso ilustra como decisões automáticas em veículos autônomos envolvem escolhas morais que não possuem respostas únicas. Alguns estudos reforçam essa discussão, como Bonnefon, Shariff e Rahwan (2016), que afirmam que os veículos autônomos, em algumas situações, terão que escolher entre duas situações ruins, como matar os seus passageiros ou alguns pedestres.

Além dos dilemas individuais citados acima, existem impactos coletivos ligados a direitos fundamentais, justiça social e democracia. UNESCO (2022) alerta que a IA traz benefícios em muitas áreas, mas se não houver a devida preocupação com questões éticas, corre o risco dessa tecnologia reproduzir vieses e discriminações do mundo real, aumentar divisões e ameaçar direitos humanos e liberdades. Muitos casos já documentados confirmam esses problemas, como o sistema de recrutamento da Amazon, que discriminava candidatas mulheres (ANGWIN et al., 2016), e o algoritmo COMPAS, usado pela justiça criminal dos Estados Unidos, acusado de atribuir maior risco de incidência a pessoas negras (DASTIN, 2018).

A seriedade do tema evidencia a necessidade de regular o desenvolvimento da inteligência artificial. Dessa forma, o AI Act da União Europeia constitui um marco regulatório ao estabelecer obrigações para aplicações de IA com base em seus riscos potenciais e nível de impacto (European Parliament, 2024). O AI Index Report 2025 observa que em 2024 houve um aumento significativo na cooperação global na governança de IA, especialmente em relação a legislação de princípios relacionadas a IA responsável e ética (GIL; PERRAULT, 2025). Esses dados sinalizam esforços globais em direção a regulamentação e a incorporação de preocupações éticas no desenvolvimento da inteligência artificial.

3.6.2 Riscos da IA

Entre os principais riscos associados à IA, destacam-se:

- **Viés e discriminação algorítmica:** sistemas treinados com dados históricos ou enviesados podem perpetuar desigualdades. A OECD (2023a) alerta que a IA pode alimentar vieses e discriminação e apresentar riscos a segurança, autonomia e a própria dignidade humana.
- **Privacidade e vazamento de dados::** modelos de IA frequentemente utilizam grandes volumes de dados e informações sensíveis, e falhas de controle ou de regulamentação podem causar a exposição ou uso indevido desses dados. Alguns casos recentes ilustram esse risco: a Clearview foi multada por armazenar ilegalmente uma base de dados contendo bilhões de imagens faciais (CORDER, 2024); o app Replika foi multado pela agência italiana de proteção de dados por violar regras destinadas a proteger os dados pessoais dos usuários (REUTERS, 2025); além disso o recente o

relatório *IBM Cost of a Data Breach 2025*, que analisou mais de 600 organizações em 17 países, mostra que 13% dessas empresas relataram brechas em modelos ou aplicações de IA, e, entre estas, 97% não tinham controles de acesso adequados (IBM, 2025).

- **Desinformação e deepfakes:** modelos generativos permitem a criação de conteúdos falsos bastante convincentes, aumentando o riscos de manipulação social. Alguns relatórios recentes destacam como a produção em massa de deepfakes e notícias falsas, têm potencial de afetar a confiança pública. O AI Index Report 2025, por exemplo, observa que a desinformação continua a representar riscos, particularmente em eleições, junto à proliferação de deepfakes. Além disso, o relatório registra que, em 2024, diversos países enfrentaram exemplos de desinformação eleitoral associada à IA, disseminada em múltiplas plataformas digitais, inclusive na eleição presidencial dos EUA (GIL; PERRAULT, 2025).
- **Automação e impacto no trabalho:** a (OECD, 2023b) alerta que milhões de empregos estão em risco serem substituídos por automações, o que exige a devida adaptação e requalificação de trabalhadores. Afirmar ainda, segundo o relatório OECD Employment Outlook 2023, que aproximadamente 27% dos empregos estão em ocupações com alto risco de automação.

3.6.3 Benefícios da IA

Entre os principais benefícios associados à IA, destacam-se:

- **Personalização de serviços:** o uso da inteligência artificial permite adaptar diagnósticos e recomendações personalizados para cada usuário em diferentes áreas, como saúde, educação e atendimento ao cliente. Relatórios internacionais ressaltam o potencial da IA para criar experiências mais individualizadas e inclusivas (UNESCO, 2022).
- **Eficiência e produtividade:** segundo a, OECD (2023a) a adoção da inteligência pode resultar em um aumento de desempenho de 20 a 40% em tarefas específicas.
- **Inclusão social:** A UNESCO (2022) ressalta a importância de garantir que os avanços em inteligência artificial e tecnologias digitais sejam orientados por princípios de inclusão de modo a beneficiar igualmente todos os países e comunidades, sem deixar ninguém para trás.
- **Apoio à decisão:** um benefício importante da IA é a capacidade de oferecer análises preditivas que auxiliam os profissionais de diferentes áreas na tomada de decisões. Nesse sentido, a World Health Organization (2021) observa que a inteligência artificial pode ser, e em alguns países já está sendo utilizada para melhorar a velocidade e precisão do diagnóstico e rastreamento de doenças, para auxiliar nos cuidados clínicos, reforçar a pesquisa em saúde e o desenvolvimento de medicamentos.

- **Escalabilidade e redução de custos:** soluções baseadas em IA permitem expandir serviços de forma rápida, sem aumentos proporcionais nos custos, o que é útil em diversos setores.

4 Relação com os conteúdos estudados

O conteúdo estudado ao longo da disciplina apresentou uma visão abrangente da inteligência artificial, deixando claro que a IA vai além do aprendizado de máquina, englobando diversas áreas e paradigmas. Essa perspectiva é afim com Russell e Norvig (2021), que descrevem a IA como um campo universal, que abrangem múltiplas abordagens e problemas. Também reforça a análise de Huyen (2024), ao mostrar como os foundation models confirmam essa diversidade.

As definições de inteligência e de IA foram tratadas em aula por meio de diferentes abordagens, como pensar e agir como humanos ou racionalmente. Essa exposição encontra fundamento nas críticas ao Teste de Turing, que mede apenas imitação, e não a inteligência em si da máquina. Russell e Norvig (2021) enfatizam essa noção, e outros autores chamam atenção para suas limitações éticas, como a dificuldade de alinhar objetivos técnicos com valores humanos.

O percurso histórico discutido em sala, desde os mitos até à conferência de Dartmouth, passando pelos invernos da IA e sistemas especialistas, também é detalhado por Russell e Norvig (2021) e no AI Index Report 2025, que descrevem a passagem do uso da inteligência artificial do laboratório para a vida cotidiana.

A descrição e o funcionamento dos agentes inteligentes, apresentados nas aulas como ciclos de percepção e ação e os diferentes tipos de agentes, são diretamente descritos por Russell e Norvig (2021), confirmando o funcionamento desses agentes na área da IA. Por fim, os debates sobre ética, riscos e benefícios, tratados em aula, encontram respaldo em diversas fontes, como a OECD, que alerta para riscos de viés e discriminação, a UNESCO que destaca ameaças a direitos fundamentais, e diversas notícias de casos reais como o sistema de recrutamento da Amazon ou o algoritmo COMPAS ilustram como esses problemas já afetam a sociedade.

Portanto, os conteúdos trabalhados durante as aulas não apenas se confirmam com outras fontes, mas também são ampliados e atualizados pelas pesquisas realizadas.

5 Conclusão

A resenha iniciou discutindo os fundamentos, a evolução histórica e o estado atual da inteligência artificial, destacando como a área se consolidou em torno dos agentes inteligentes, e também como se expandiu com os avanços do deep learning e do foundation models. Posteriormente, foram analisados os riscos associados ao uso da inteligência artificial, como vieses, privacidade e desinformação, e também os benefícios potenciais, como personalização de serviços, eficiência, apoio à decisão e inclusão social.

A importância do tema é evidente, já que a IA deixou de ser apenas uma área técnica e tornou-se uma questão estratégica para a ciência, economia e sociedade, exigindo regulação, preocupações éticas e participação multidisciplinar.

Como reflexão futura, acredito que o grande desafio da inteligência artificial será garantir que seus objetivos estejam de acordo com os valores humanos, equilibrando as inovações com responsabilidade. Também será importante investir em pesquisas que tornem a tecnologia mais inclusiva e justa, para que os avanços ajudem a sociedade em vez de aumentar desigualdades.

Referências

- ANGWIN, J. et al. *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*. 2016. ProPublica. Acesso em: 22 set. 2025. Disponível em: <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>.
- CORDER, M. *Clearview AI fined in Netherlands over facial recognition privacy violations*. 2024. Acesso em: 21 set. 2025. Disponível em: <<https://apnews.com/article/clearview-ai-facial-recognition-privacy-fine-netherlands-a1ac33c15d561d37a923b6c382f48ab4>>.
- DASTIN, J. *Amazon scraps secret AI recruiting tool that showed bias against women*. 2018. Reuters. Acesso em: 22 set. 2025. Disponível em: <<https://www.reuters.com/article/amazon-ai-recruiting-idUSKCN1MK08G/>>.
- European Parliament. *Artificial Intelligence Act: MEPs adopt landmark law*. 2024. Acesso em: 22 set. 2025. Disponível em: <<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>>.
- GIL, Y.; PERRAULT, R. (Ed.). *Artificial Intelligence Index Report 2025*. 2025. Stanford University. Acesso em: 18 set. 2025. Disponível em: <<https://aiindex.stanford.edu/report/>>.
- HUYEN, C. *AI Engineering: Building Applications with Foundation Models*. [S.l.]: O'Reilly Media, 2024.
- IBM. *IBM Report: 13% of organizations reported breaches of AI models or applications, 97% of which reported lacking proper AI access controls*. 2025. Acesso em: 22 set. 2025. Disponível em: <<https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls>>.
- MIT Media Lab. *Moral Machine*. 2016. Acesso em: 20 set. 2025. Disponível em: <<https://www.moralmachine.net/>>.
- OECD. *Artificial Intelligence – Policy Issues*. 2023. Acesso em: 22 set. 2025. Disponível em: <<https://www.oecd.org/en/topics/policy-issues/artificial-intelligence.html/#key-messages>>.
- OECD. *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. 2023. Acesso em: 22 set. 2025. Disponível em: <https://www.oecd.org/en/publications/oecd-employment-outlook-2023/_08785bba-en.html>.
- OPPY, G.; DOWE, D. *The Turing Test*. 2021. In: Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.). Acesso em: 19 set. 2025. Disponível em: <<https://plato.stanford.edu/entries/turing-test/>>.
- REUTERS. *Italy's data watchdog fines AI company Replika's developer \$5.6 million*. 2025. Acesso em: 21 set. 2025. Disponível em: <<https://www.reuters.com/sustainability/boards-policy-regulation/italys-data-watchdog-fines-ai-company-replikas-developer-56-million-2025-05-19>>.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed. New Jersey: Pearson, 2021.
- Stanford University. *Ancient myths reveal early fantasies about artificial life*. 2019. Acesso em: 20 set. 2025. Disponível em: <<https://news.stanford.edu/stories/2019/02/ancient-myths-reveal-early-fantasies-artificial-life>>.

UNESCO. *Artificial Intelligence*. 2022. Acesso em: 20 set. 2025. Disponível em: <<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>>.

World Health Organization. *WHO issues first global report on AI in health and six guiding principles for its design and use*. 2021. Acesso em: 22 set. 2025. Disponível em: <<https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use>>.