

ÉTICA NA INTELIGÊNCIA ARTIFICIAL: Desenvolvimento Responsável de Sistemas Inteligentes

ETHICS IN ARTIFICIAL INTELLIGENCE: Responsible Development of Intelligent Systems

Juliano Augusto Anselmo Gema – juliano-gema@hotmail.com
Faculdade de Tecnologia de Taquaritinga (Fatec) – Taquaritinga – São Paulo – Brasil

João de Lucca Filho – joaodelucca@terra.com.br
Faculdade de Tecnologia de Taquaritinga (Fatec) – Taquaritinga – São Paulo – Brasil

DOI: 10.31510/inf.v21i1.1954

Data de submissão: 14/04/2024

Data do aceite: 10/03/2024

Data da publicação: 20/06/2024

RESUMO

Este trabalho aborda a importância crítica da ética na evolução e implementação da inteligência artificial (IA), destacando os importantes desafios éticos significativos que surgem da integração em diferentes aspectos da vida humana. Ao explorar os fundamentos éticos necessários para um desenvolvimento responsável de sistemas inteligentes, garante-se em garantir que a IA seja guiada por princípios éticos fundamentais, como justiça, transparência e integridade. Através de uma revisão sistemática da literatura, análise de estudos de caso e avaliação de princípios éticos estabelecidos por organizações líderes, esta pesquisa fornece orientação prática para a implementação de melhores práticas. Os resultados destacam a necessidade urgente de frameworks éticos para orientar o desenvolvimento da IA, demonstrando que a adoção de medidas transparentes e justas não só minimiza o risco de preconceitos e violações da privacidade, mas também promove maior confiança e aceitação social dos sistemas de IA. Por fim, conclui-se que a integração de considerações éticas desde o início do processo de desenvolvimento da IA é essencial para construir um futuro tecnológico que seja tanto inovador e consistentes com os valores humanos fundamentais.

Palavras-chave: Ética em Inteligência Artificial. Transparência. Responsabilidade. Viés Algorítmico. Privacidade.

ABSTRACT

This work addresses the critical importance of ethics in the evolution and implementation of artificial intelligence (AI), highlighting the significant ethical challenges that arise from integration into different aspects of human life. By exploring the ethical foundations necessary for responsible development of intelligent systems, we ensure that AI is guided by fundamental ethical principles such as fairness, transparency and integrity. Through a systematic literature review, case study analysis, and assessment of ethical principles established by leading organizations, this research provides practical guidance for implementing best practices. The results highlight the urgent need for ethical frameworks to guide the development of AI,

demonstrating that adopting transparent and fair measures not only minimizes the risk of bias and privacy violations, but also promotes greater trust and social acceptance of AI systems. Finally, it is concluded that integrating ethical considerations from the beginning of the AI development process is essential to building a technological future that is both innovative and consistent with fundamental human values.

Keywords: Ethics in Artificial Intelligence. Transparency. Responsibility. Algorithmic Bias. Privacy.

1 INTRODUÇÃO

Na era da transformação digital, a inteligência artificial (IA) emerge como uma força revolucionária com o potencial para remodelar inúmeras aspectos da vida humana. Da medicina à educação e à tomada de decisões críticas, a rápida expansão da IA promete inovação sem precedentes. No entanto, a crescente integração da IA com nosso cotidiano levanta questões prementes sobre a ética na tecnologia.

O desenvolvimento de sistemas inteligentes apresenta-nos desafios éticos importantes destas tecnologias e de como respeitar princípios éticos básicos como a justiça, a transparência e a responsabilização. Este trabalho centra-se na questão de como garantir o desenvolvimento responsável de sistemas inteligentes que respeitem estes princípios éticos, explora os fundamentos éticos necessários e analisa os desafios que surgem éticas específicas.

O objetivo é contribuir para o campo da ética em tecnológica, fornecendo diretrizes para a prática ética da IA. A relevância deste estudo reside na sua capacidade de influenciar positivamente a trajetória futura da IA, garantindo que o seu desenvolvimento seja guiado por rigorosas considerações éticas, contribuindo assim para a criação de uma sociedade mais justa e igualitária.

Através de uma abordagem metodológica que inclui análise da literatura existente, estudos de caso e diretrizes éticas propostas por organizações líderes, este trabalho procura fornecer insights.

Através de uma abordagem metodológica que inclui análise da literatura existente, em livros, bem como levantamento de artigos e periódicos, trabalhos acadêmicos, estudos de caso e diretrizes éticas propostas por organizações líderes, este trabalho procura fornecer insights. Há valor em como a ética pode ser integrada no coração do desenvolvimento da IA.

2 FUNDAMENTOS ÉTICOS EM IA

A emergência da inteligência artificial (IA) como uma força transformadora na sociedade moderna exige uma revisão ética rigorosa. A base teórica deste trabalho é baseada em uma extensa revisão da literatura existente sobre ética em IA, proeminentes na área e a conselhos de organizações influentes.

2.1 Transparência

A transparência é amplamente reconhecida como um pilar essencial para da ética da IA. Coeckelbergh (2020) argumenta que a transparência no processo de tomada de decisão dos sistemas de IA é fundamental para estabelecer a confiança e responsabilização, sugerindo que “a transparência não é apenas uma questão técnica, mas também uma questão ética profunda” (COECKELBERGH, 2020, p. 112).

Esta visão é reforçada pela European Commission's High-Level Expert Group on AI (2019), que postula que a transparência deve abranger tanto os aspectos técnicos como os impactos sociais dos sistemas de IA.

2.2 Justiça

A questão da justiça na IA envolve com a prevenção de preconceitos e discriminação, garantindo que os sistemas de IA funcionam de forma justa. Russell e Norvig (2013) consideram como os algoritmos podem involuntariamente ou mesmo exacerbar as desigualdades sociais, enfatizando a importância de métodos de desenvolver abordagens que busquem evitar preconceitos.

O'Neil (2016) complementa esta discussão com uma análise de como “armas de destruição matemática” podem amplificar a desigualdade, destacando a necessidade de uma abordagem ética rigorosa no design de algoritmos.

2.3 Responsabilidade

A responsabilização envolve a atribuição de responsabilidade ética e legal aos criadores e utilizadores de sistemas de IA. Bostrom e Yudkowsky (2014) exploram as complexidades da atribuição de responsabilidades em sistemas autônomos, argumentando que “a criação de sistemas de IA com a capacidade de tomar decisões autônomas eleva questões significativas sobre a responsabilidade por ações realizadas por estes sistemas” (BOSTROM; YUDKOWSKY, 2014, p. 320).

Esta discussão é essencial para compreender como manter a responsabilidade no contexto de sistemas cada vez mais autônomos.

3 PROCEDIMENTOS METODOLÓGICOS

Para conduzir esta revisão sobre os fundamentos éticos na inteligência artificial (IA), foram realizadas buscas sistemáticas em bases de dados eletrônicas renomadas, como SCIELO, PubMed e Google Acadêmico. Também foram consultados sites especializados e obras especializadas em ética tecnológica e inteligência artificial.

As pesquisas concentraram a cobertura cronológica de estudo realizado nos últimos 5 anos. Alguns trabalhos fundamentais relacionados à ética em IA foram meticulosamente selecionados para a análise.

Para a busca bibliográfica, foram utilizadas as palavras-chave: “ética em inteligência artificial”, “desenvolvimento responsável de IA”, “princípios éticos em sistemas inteligentes”, “transparência em IA”, “justiça algorítmica”, e “responsabilidade em sistemas de IA”.

4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos na pesquisa e os discute à luz da teoria discutida, explorando a relação entre os resultados e as teorias existentes sobre a ética na inteligência artificial (IA).

Os resultados são analisados criticamente, tendo em conta desafios éticos específicos, a transparência e interpretabilidade, responsabilidades legais e sociais, impacto nas decisões e na sociedade, bem como a perspectiva empresarial futura do desenvolvimento ético na IA.

4.1 Desafios Éticos Específicos em IA

Uma investigação sobre os desafios éticos específicos na inteligência artificial (IA) revelou que o preconceito algorítmico, a privacidade, a segurança e a responsabilidade são preocupações fundamentais no desenvolvimento e implantação de sistemas inteligentes. Ao examinar casos específicos da vida real, este estudo destaca a complexidade destes desafios e avalia abordagens e soluções testadas.

4.1.1 Viés Algorítmico: Caso do Sistema de Recrutamento da Amazon

Um dos exemplos mais emblemáticos do viés algorítmico ocorreu com o sistema de recrutamento baseado em IA da Amazon, que foi descartado depois que descobriu que favorecia candidatos masculinos para cargos técnicos.

Este viés se deve ao sistema de formação com currículos enviados à empresa com mais de 10 anos, principalmente por homens. Este caso destaca a importância de uma abordagem cuidadosa na seleção e processamento de dados de treinamento para sistemas de IA, destacando como preconceitos inconscientes podem ser inadvertidamente codificados em algoritmos.

4.1.2 Privacidade: Caso do Reconhecimento Facial em Espaços Públicos

A implantação de sistemas de reconhecimento facial em espaços públicos por governos e organizações privadas provocou um amplo debate sobre privacidade e vigilância. Um caso infame ocorreu em uma grande cidade onde a polícia começou a usar tecnologia de reconhecimento facial para identificar suspeitos em tempo real.

A falta de transparência sobre a forma como os dados são utilizados e armazenados levantou preocupações significativas sobre a privacidade dos cidadãos e o potencial de abuso. Este caso destaca a necessidade de diretrizes claras sobre coleta, utilização e armazenamento de dados pessoais, bem como de consentimento explícito por parte daqueles que estão a ser monitorizados.

4.1.3 Segurança: Incidente com Carros Autônomos

O desenvolvimento de veículos autônomos revelou sérios desafios de segurança, ilustrados pelo primeiro acidente fatal envolvendo um veículo autônomo e um pedestre. A investigação subsequente revelou falhas na forma como o sistema de IA do veículo detectava e respondia a obstáculos inesperados.

Este caso ilustra a importância do desenvolvimento de sistemas de IA que sejam robustos e seguros, destaca a necessidade de testes rigorosos em diferentes cenários antes da implantação generalizada.

4.1.4 Responsabilidade: Uso de IA na Tomada de Decisões Judiciais

A utilização de sistemas de IA para apoiar a tomada de decisões judiciais levantou questões sobre a responsabilização, especialmente nos casos em que a IA impõe penas mais severas a determinados grupos demográficos.

Um caso específico envolveu software de “risco de reincidência” utilizado pelos juízes para informar as suas decisões de sentença, que foi criticado por perpetuar preconceitos existentes no sistema judicial. Este caso destaca a complexidade da atribuição de responsabilidades nos sistemas de IA e a necessidade de garantir que estes sistemas sejam transparentes, justos e auditáveis.

4.2 Transparência e Interpretabilidade

A transparência e a interpretabilidade dos sistemas de inteligência artificial (IA) são essenciais para garantir que os usuários compreendem como as decisões são tomadas, o que gera confiança e responsabilização. A análise dos dados coletados neste estudo destaca a crescente importância destes conceitos no desenvolvimento ético da IA, em linha com as discussões teóricas atuais.

Coeckelbergh (2020) argumenta que a transparência na IA não deve ser vista apenas como uma questão técnica, mas também como um imperativo ético que requer esclarecer como os algoritmos operam e como suas decisões afetam indivíduos e a sociedade. Este ponto é essencial para evitar os riscos de mal-entendidos ou desconfianças que podem surgir de sistemas considerados “caixas-pretas”. A necessidade de sistemas inteligentes que possam ser interpretados e compreendidos por não especialistas também é destacada, sugerindo um movimento no sentido de desenvolver métodos para os usuários questionarem e compreender que as decisões são tomadas pela IA.

A European Commission's High-Level Expert Group on AI (2019) enfatiza a importância da transparência na criação de uma IA confiável, incluindo a clareza dos processos de tomada de decisão e a possibilidade de auditabilidade. Esta orientação sugere que, além do desenvolvimento de sistemas transparentes, há necessidade de implementar estruturas que permitam que decisões automatizadas sejam revistas e desafiadas, promovendo assim um ciclo virtuoso de confiança e melhoria contínua.

No entanto, os desafios na implementação da transparência e da interpretabilidade na prática são enormes. Russell e Norvig (2013) discutem as limitações técnicas atuais que dificultam a criação de sistemas de IA totalmente transparentes e compreensíveis, especialmente em modelos complexos, como redes neurais profundas. Enfatizam a necessidade de investigação contínua para desenvolver novas técnicas que possam tornar a IA mais acessível e compreensível para todos os utilizadores.

Os resultados deste estudo confirmam a teoria existente, destacando tanto a importância e os desafios de promover a transparência e a interpretabilidade da IA. Sugerem também sugerem que, embora haja um consenso crescente sobre a necessidade destes princípios, a sua aplicação eficaz ainda enfrenta obstáculos significativos, reforçando a necessidade de abordagens inovadoras criadas para equilibrar a complexidade técnica com a acessibilidade e a facilidade de compreensão.

4.3 Impacto nas Decisões e na Sociedade

A implantação de sistemas de IA terá um impacto profundo nas decisões sociais e individuais, moldando setores importantes como a saúde, a educação e a justiça. Estes sistemas oferecem o potencial para melhorar significativamente a eficiência e a objetividade, por exemplo no diagnóstico médico precoce e na aprendizagem personalizada. No entanto, os resultados deste estudo também destacam que a IA pode amplificar involuntariamente as desigualdades existentes e marginalizar grupos vulneráveis.

A análise mostra que o impacto social da IA está intrinsecamente relacionado com a sua programação e os dados sobre os quais é treinada. Por exemplo, os sistemas de IA utilizados no recrutamento podem perpetuar preconceitos contra candidatos de determinadas origens se os dados históricos refletirem práticas de contratação discriminatórias. Da mesma forma, os algoritmos utilizados no sistema judiciário podem reproduzir desigualdades sistêmicas, levantando questões sobre justiça e equidade.

Este cenário destaca a necessidade crítica de uma abordagem ética ao desenvolvimento e implantação da IA, que leva em conta não só os benefícios potenciais, mas também os riscos de exacerbar a desigualdade social. A colaboração entre desenvolvedores de IA, cientistas sociais, especialistas em ética e comunidades afetadas é essencial para garantir que os sistemas de IA sejam concebidos, projetados e implementados de forma justa e equitativa, promovendo assim o bem-estar de todos os membros da sociedade.

4.4 Perspectivas para o Futuro

Olhando para o futuro, a integração da ética no desenvolvimento da IA enfrentará desafios novos e oportunidades. A evolução contínua da tecnologia e a crescente adoção de sistemas autônomos realçam a necessidade de abordagens dinâmicas e adaptativas à ética na IA. Uma das principais áreas de foco será o desenvolvimento de sistemas de IA explicáveis, que não só operem de forma transparente, mas também tomem decisões que os utilizadores possam facilmente compreender e justificar.

Além disso, a governança global da IA surge como um domínio crítico, exigindo cooperação internacional para estabelecer padrões e regulamentos éticos. A diversidade cultural e as diferenças nos sistemas jurídicos colocam desafios especiais na criação de um consenso global sobre a ética da IA. No entanto, a cooperação internacional é vital para evitar um “nivelamento por baixo” nos padrões éticos, garantindo que o desenvolvimento da IA seja guiados por valores compartilhados de dignidade humana e justiça.

Finalmente fim, a educação ética em IA torna-se cada vez mais importante, tanto para os desenvolvedores como para o público. Desenvolver uma compreensão ampla dos dilemas éticos associados à IA e cultivar o pensamento crítico sobre a sua utilização é essencial para promover uma sociedade informada e capacitada. Os programas de educação em ciência, tecnologia, engenharia e matemática (STEM) e em humanidades devem incorporar a ética da IA, preparando a próxima geração para navegar de forma responsável no futuro tecnológico.

Estas perspectivas futuras destacam a importância de uma abordagem proativa e colaborativa para enfrentar os desafios éticos da IA, garantindo que a inovação tecnológica avança de forma a respeitar os direitos humanos e promover uma sociedade justa e igualitária.

4.5 Colaboração Internacional e Governança Global da IA

A governança global eficaz da inteligência artificial é um crucial desafio que requer esforços de colaboração internacional. As iniciativas de organizações internacionais desempenham um papel importante na definição das políticas nacionais de IA, estabelecendo princípios éticos e diretrizes de governança que transcendem fronteiras.

4.5.1 Iniciativas Globais e Seu Impacto

OCDE e Seus Princípios sobre IA: Em 2019, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) adotou os Princípios sobre IA, que enfatizam a inclusão, a sustentabilidade, a transparência e a responsabilidade. Estes princípios sustentaram uma série políticas nacionais de IA. Por exemplo, o Japão incorporou estes princípios na sua “Estratégia de Sociedade IA”, que se concentra no desenvolvimento de IA que beneficie a sociedade e garanta a segurança e a privacidade dos dados. Além disso, a França e o Canadá lançaram o Global Partnership on AI (GPAI), uma iniciativa para promover o desenvolvimento responsável da IA de acordo com os princípios da OCDE.

UNESCO e Ética da IA: A UNESCO está liderando esforços para estabelecer as primeiras recomendações globais sobre a ética da IA, com o objetivo de orientar os seus membros na formulação de políticas que garantam o desenvolvimento ético da IA. Esta recomendação está molda as discussões nacionais sobre a ética da IA, como visto no Brasil, onde a elaboração de uma política nacional de IA leva em conta as diretrizes da UNESCO para promover uma IA que respeite os direitos humanos e a diversidade cultural.

4.5.2 Desafios e Oportunidades

A implementação destas diretrizes globais enfrenta desafios, incluindo a adaptação aos contextos culturais e jurídicos locais. No entanto, isto oferece oportunidades únicas para estabelecer uma base comum sobre valores éticos em IA. A cooperação internacional, como exemplificada pelo GPAI, proporciona um fórum para os países compartilharem melhores práticas, investigação, pesquisas e políticas para enfrentar os desafios comuns de governança da IA.

4.5.3 Chamado à Ação

Para uma governança global bem-sucedida da IA, é imperativo o envolvimento contínuo de todos os *stakeholders*: governos, setor privado, mundo acadêmico e sociedade civil. Iniciativas como as da OCDE e UNESCO proporcionam o quadro de referência para essa cooperação, mas a sua eficácia depende da implementação e da participação ativa a nível nacional e local. É portanto, importante que os países adotem estas recomendações, adaptando-as conforme necessário para refletir as suas próprias necessidades e valores, mantendo-se ao mesmo tempo comprometidos com princípios éticos universais.

4.6 Caminhos para uma IA ética e responsável

4.6.1 Soluções Inovadoras e Iniciativas Futuras

A integração eficaz da ética no desenvolvimento da IA não é apenas desejável, mas necessária para enfrentar os desafios atuais e futuros. Como apontam por Floridi e Cowls (2019), frameworks éticos como o proposto pela European Commission's High-Level Expert Group on AI, oferecem “orientações fundamentais para a ação ética” no desenvolvimento da IA, promovendo práticas que respeitam a dignidade humana, a liberdade e a diversidade.

Iniciativas como IA explicável (XAI) representam esforços importantes para aumentar a transparência e a compreensibilidade da IA, o que é essencial para a confiança e a cooperação entre humanos e sistemas inteligentes (ARICLETA et al., 2020).

4.6.2 A Importância da Educação e Formação em Ética

A educação em ética é essencial para preparar os desenvolvedores e outros stakeholders para os desafios éticos da IA. Segundo Mittelstadt (2019), integrar a ética na formação de cientistas e engenheiros de dados pode “fomentar uma cultura de responsabilidade ética”, necessária para o desenvolvimento de tecnologia que beneficie toda a sociedade mundial.

Programas educacionais e workshops, como os oferecidos pelo AI Ethics Lab, demonstram um compromisso contínuo com a formação ética, equipando os profissionais com

as ferramentas necessárias para avaliar criticamente o impacto total do seu trabalho (CONITZER, BAUM, 2018).

4.6.3 Para Além: Uma Visão Colaborativa e Multidisciplinar

A colaboração e o diálogo entre disciplinas são fundamentais para uma ética forte na IA. Conforme apontado por Rahwan et al. (2019), enfrentar os desafios éticos da IA requer “uma abordagem multidisciplinar que reúna acadêmicos, desenvolvedores, reguladores e o público em geral”.

Esta colaboração pode promover uma compreensão mais rica do impacto social da IA e facilitar a criação de tecnologias mais consistentes com valores éticos compartilhados. Fóruns de discussão abertos e iniciativas como o AI4Good Global Summit exemplificam os esforços para envolver diversas vozes sobre a direção futura da IA (VINUESA et al., 2020).

5 CONCLUSÃO

Embora os desafios éticos associados à IA sejam vastos e complexos, as soluções e iniciativas atuais demonstram um compromisso crescente com o desenvolvimento responsável destes sistemas. A educação e a formação em ética desempenham um papel fundamental na construção de uma base sólida para este esforço, preparando todos os participantes para navegarem no difícil terreno da IA com integridade e visão de futuro.

Por fim, ao se adotar estas soluções e continuar a procurar inovações éticas, pode-se esperar uma era em que a IA não só melhore as nossas capacidades, mas também reflita os nossos valores mais elevados, promovendo uma sociedade mais justa e equitativa.

REFERÊNCIAS

- ARICLETA, A. B.; et al. *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information Fusion, v. 58, p. 82-115, 2020.
- BOSTROM, N.; YUDKOWSKY, E. *The Ethics of Artificial Intelligence*. In: FRANKISH, Keith; RAMSEY, William M. (Eds.). *Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 2014. p. 316-334.
- COECKELBERGH, M. *AI Ethics*. Cambridge, MA: The MIT Press, 2020.
- CONITZER, V.; BAUM, S. D. *Moral Decision Making Frameworks for Artificial Intelligence*. In: AAAI Conference on Artificial Intelligence. 2018.

EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. *Ethics Guidelines for Trustworthy AI*. Bruxelas: European Commission, 2019.

FLORIDI, L.; COWLS, J. *A Unified Framework of Five Principles for AI in Society*. Harvard Data Science Review. 2019.

MITTELSTADT, B. *Principles alone cannot guarantee ethical AI*. **Nature Machine Intelligence**, v. 1, n. 11, p. 501-507, 2019.

O'NEIL, C. **Armas de Destruição Matemática: Como o Big Data Aumenta a Desigualdade e Ameaça a Democracia**. Rio de Janeiro: Zahar, 2016.

RAHWAN, I. et al. *Machine behaviour*. Nature, v. 568, p. 477-486, 2019.

RUSSELL, S. J.; NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.

VINUESA, R.; et al. *The role of artificial intelligence in achieving the Sustainable Development Goals*. Nature Communications, v. 11, n. 1, 2020.