



Universidade Federal de Ouro Preto
Escola de Minas
CECAU - Colegiado do Curso de
Engenharia de Controle e Automação



Sara Gonçalves Ferreira Pereira

**O verão da Inteligência Artificial: um estudo sobre ética e alcance das
ferramentas computacionais na Era do *big data***

Monografia de Graduação

Ouro Preto, 2023

Sara Gonçalves Ferreira Pereira

**O verão da Inteligência Artificial: um estudo sobre ética e
alcance das ferramentas computacionais na Era do *big
data***

Trabalho apresentado ao Colegiado do Curso
de Engenharia de Controle e Automação da
Universidade Federal de Ouro Preto como
parte dos requisitos para a obtenção do Grau
de Engenharia de Controle e Automação.

Universidade Federal de Ouro Preto

Orientador: Prof. Dr. Danny Augusto Vieira Tonidandel

Ouro Preto

2023



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA CONTROLE E
AUTOMACAO



FOLHA DE APROVAÇÃO

Sara Gonçalves Ferreira Pereira

**O verão da inteligência artificial:
um estudo sobre ética e alcance das ferramentas computacionais na era do big data**

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia de Controle e Automação

Aprovada em 26 de outubro de 2022.

Membros da banca

[Doutor] - Danny Augusto Vieira Tonidandel - Orientador (Universidade Federal de Ouro Preto)

[Doutora] - Adrielle de Carvalho Santana - (Universidade Federal de Ouro Preto)

[Doutor] - Paulo Marcos de Barros Monteiro - (Universidade Federal de Ouro Preto)

Danny Augusto Vieira Tonidandel, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 26/10/2023.



Documento assinado eletronicamente por **Danny Augusto Vieira Tonidandel, PROFESSOR DE MAGISTERIO SUPERIOR**, em 26/10/2023, às 18:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0612482** e o código CRC **93B1370A**.

Agradecimentos

Em primeiro lugar, agradeço a Deus por abençoar minha vida e por me guiar até aqui.

Aos meus pais, José Carlos e Denise, por serem meus pilares e exemplo de amor e dedicação constante. Aos meus irmãos, Samira e Saulo, pelo carinho e por me impulsionarem a sonhar. À minha família, por todo o apoio e companheirismo.

À comunidade da UFOP e à equipe docente pelos valiosos ensinamentos. Em especial, ao meu orientador, prof. Danny Tonidandel, por todo o suporte.

Por fim, agradeço a todos aqueles que forneceram contribuições valiosas, seja por meio de *insights*, recursos ou suporte prático, para a realização desta pesquisa .

“Somos a memória que temos e a responsabilidade que assumimos. Sem memória não existimos, sem responsabilidade talvez não mereçamos existir” (José Saramago)

Resumo

No chamado “verão da inteligência artificial” o próprio imaginário coletivo vem sendo modulado a partir dos avanços das pesquisas em inteligência computacional, muitas vezes divulgada por meio de um viés ficcional e fantasioso, com pouca ou quase nenhuma conexão com a realidade. Neste aspecto, a compreensão das implicações éticas geradas pelas novas invenções se apresenta como vasto território a ser explorado. Não obstante a onipresença da conexão em rede e do turbilhão de dados digitais gerados ininterruptamente, há poucos trabalhos que visam pavimentar os caminhos para um progresso robusto da inteligência de máquina, e sua aplicação responsável. Objetiva-se, portanto, levantar questões concernentes aos desafios éticos ligados ao *Big Data* e, em especial, aos sistemas de Inteligência Artificial (IA). Para isto, como aporte teórico, foram utilizadas algumas reflexões do Filósofo italiano e professor da universidade de Oxford, Luciano Floridi, que tratam sobretudo da identificação e classificação dos riscos envolvidos na produção, utilização e distribuição das ferramentas de IA. Devido à diversidade de abordagens técnicas, nem sempre é possível enquadrá-los em uma categoria específica. Portanto, o presente estudo busca lançar luz sobre algumas dessas questões, de forma a se constituir em um guia introdutório para os estudos de governança digital, com vistas a garantir um progresso eticamente responsável da pesquisa e aplicação da Inteligência Artificial, bem como suas imensas implicações para o futuro da humanidade.

Palavras-chaves: Inteligência Artificial. *big data*. Ética. Governança. Privacidade. Transparência. Responsabilidade. Tomada de Decisão.

Abstract

In the so-called “summer of artificial intelligence” the collective imagination itself has been modulated based on advances in research in computational intelligence, often disseminated through a fictional and fantasy bias, with little or almost no connection with reality. In this aspect, understanding the ethical implications generated by new inventions presents itself as a vast territory to be explored. Despite the ubiquity of networking and the whirlwind of digital data generated uninterruptedly, there is little work that aims to pave the way for robust progress in machine intelligence and its responsible application. The aim, therefore, is to raise questions concerning the ethical challenges linked to *Big Data* and, in particular, to Artificial Intelligence (AI) systems. For this, as a theoretical contribution, some reflections from the Italian philosopher and professor at the University of Oxford, Luciano Floridi, were used, which mainly deal with the identification and classification of risks involved in the production, use and distribution of AI tools. Due to the diversity of technical approaches, it is not always possible to classify them into a specific category. Therefore, the present study seeks to shed light on some of these issues, in order to constitute an introductory guide for digital governance studies, with a view to ensuring ethically responsible progress in the research and application of Artificial Intelligence, as well as its immense implications for the future of humanity.

Key-words: Artificial Intelligence. *big data*. Ethics. Governance. Privacy. Transparency. Responsibility. Decision Making.

Lista de ilustrações

Figura 1 – Pesquisa em Inteligencia Artificial no Brasil. Fonte: GOOGLE... (2023).	9
Figura 2 – Agente de utilidade baseado em modelos. Fonte: Russel (2013).	13
Figura 3 – Evolução industrial.	21
Figura 4 – Agente de utilidade baseado em modelos. Fonte: Baduge et al. (2022).	22
Figura 5 – Pirâmide de riscos. Fonte: Madiega (2021).	32
Figura 6 – Matriz de risco. Fonte: Krafft, Zweig e König (2022).	34
Figura 7 – Conceito XAI adaptado de: Gunning (2016).	37
Figura 8 – Comparação XAI. Fonte: Gunning (2016).	38
Figura 9 – Funções e responsabilidades durante auditorias independentes. Fonte: Mokander et al. (2021)	39
Figura 10 – Tabela resumo do Código de Ética e Conduta Profissional. Fonte: Adaptado da ACM Piteira, Aparicio e Costa (2019).	45

Sumário

1	INTRODUÇÃO	9
1.1	Objetivos	10
1.2	Metodologia	11
1.3	Estrutura	11
2	O QUE HÁ DE REAL NA INTELIGÊNCIA ARTIFICIAL?	12
2.1	O que há de artificial na IA?	12
2.2	Responsabilidade na utilização da inteligência de máquina	14
2.3	Uma Filosofia para a Informação	16
2.4	IA na Engenharia	20
3	O VERÃO DA INTELIGÊNCIA ARTIFICIAL	24
3.1	O desafio da Automação onipresente	24
3.2	IA no ambiente de trabalho	26
3.3	Estratégias de Governança Ética na Era da IA	38
3.4	Trilhando o Caminho Ético	43
4	CONSIDERAÇÕES FINAIS	47
	Conclusão	50
	Referências	52

1 Introdução

Diante do impressionante avanço tecnológico nas duas primeiras décadas do século XXI, o desenvolvimento da Inteligência Artificial (IA) tem sido meteórico. O que antes era confinado ao mundo da ficção científica, hoje parece ter se tornado uma parte tangível e essencial do cotidiano. Essa transformação é palpável, com a aceleração das capacidades da IA atingindo alturas sem precedentes nos últimos anos. Para se compreender a evolução do interesse público em relação à Inteligência Artificial (IA), uma análise foi conduzida utilizando a ferramenta *Google Trends*. Tal análise foi baseada na busca por termos como “Inteligência Artificial” e sua sigla *IA* no território brasileiro e global, abrangendo o período compreendido de aproximadamente duas décadas, entre 01 de janeiro de 2004 a junho de 2023.

Os resultados dessa análise (figura 1) ilustram uma ascensão notável e aparentemente exponencial no interesse mundial pelo campo de estudo da IA ao longo das últimas décadas, uma vez que a pesquisa em IA teve início em meados do século passado, em paralelo à invenção do computador digital após o fim da segunda guerra mundial. Observa-se uma tendência semelhante nas buscas relacionadas, como *AI chatbots* e *character.ia*, indicando uma crescente curiosidade em tópicos correlatos. É importante ressaltar que essa análise se baseia nas tendências de busca online e, portanto, reflete o interesse manifesto da comunidade virtual em torno do assunto.

Esses dados preliminares oferecem um panorama valioso para compreender a disseminação e a relevância crescentes da IA na sociedade contemporânea, ao fornecer um panorama propício ao aprofundamento sobre os impactos e as implicações da IA, bem como explorar os desafios éticos e técnicos que a acompanham.



Figura 1 – Pesquisa em Inteligencia Artificial no Brasil. Fonte: [GOOGLE... \(2023\)](#).

A análise de tendências de interesse em Inteligência Artificial contribui embasa as

próximas seções deste estudo, nas quais serão explorados em detalhes os aspectos relevantes da evolução e aplicação da IA em diversas áreas. Dessa forma, percebe-se que é crescente, e cada vez mais frequente, o emprego da inteligência artificial em múltiplos domínios do cotidiano, desde ambientes corporativos e industriais até a utilização de ferramentas pessoais como a *Alexa* e o *ChatGPT*, destacando a necessidade de investigar os desafios éticos intrínsecos a essa expansão substancial.

Também é importante sustentar que da IA na engenharia seja abordada de maneira equilibrada, compreendendo tanto os aspectos técnicos quanto suas implicações sociais e éticas. Dessa forma, nota-se que a compreensão dos desafios técnicos enfrentados na implementação da IA em engenharia é fundamental para aprimorar as abordagens e maximizar os resultados.

Portanto, a pesquisa busca sugerir medidas a serem tomadas para que o progresso tecnológico em termos de IA ocorra de maneira sustentável e benéfica, para toda sociedade. Além disso, espera-se contribuir para a formação de profissionais capacitados a aplicar IA de maneira eticamente responsável, moldando um futuro tecnológico mais promissor e menos distópico.

1.1 Objetivos

Geral

Apresentar possíveis caminhos para um desenvolvimento seguro da inteligência artificial e do aprendizado de máquina que conservem, em sua essência, a dignidade humana.

Objetivos Específicos

- Explorar alguns dos princípios éticos já propostos no contexto da inteligência artificial e sistemas automatizados;
- Avaliar os impactos sociais da Inteligência Artificial, considerando pesquisas e estudos de especialistas focados na empregabilidade, privacidade de dados e o uso de ferramentas;
- Propor orientações e sugestões para uma implementação e comportamento responsável de engenheiros e desenvolvedores de *softwares* com IA, enfatizando a importância da transparência.

1.2 Metodologia

Para que tais objetivos seja alcançados, serão utilizadas como referência estrutural conceituações do Filósofo italiano e professor da Universidade de Oxford, Luciano Floridi, que tem focado suas pesquisas nos sistemas de inteligência artificial, além de outros pesquisadores, no intuito de ressaltar a criação diretrizes mais robustas para o desenvolvimento da Inteligência Artificial, sobretudo no que concerne às práticas da Engenharia de Controle e Automação e sua influência na sociedade e na indústria.

A metodologia adotada neste trabalho é uma combinação de pesquisa qualitativa e descritiva. Essa abordagem permite uma análise relativamente aprofundada dos desafios éticos e responsabilidades associadas à evolução e uso da inteligência artificial e na sua aplicação em sistemas de controle e automação.

A pesquisa qualitativa é utilizada para explorar as implicações éticas da inteligência artificial na sociedade e na tomada de decisões. Isso envolve a análise de dados não numéricos, como discursos, trabalhos e visões de especialistas em ética e IA. A pesquisa descritiva é aplicada para analisar de as características da expansão da inteligência artificial em diversos domínios. Isto envolverá a coleta de dados sobre como a IA está sendo aplicada em ambientes corporativos, industriais e pessoais.

No entanto, reconhece-se as possíveis limitações da pesquisa, bem como os desafios enfrentados, como a disponibilidade de dados e a natureza subjetiva das opiniões. Considera-se, assim, a necessidade de uma abordagem mais flexível para capturar nuances éticas e identificar áreas onde será possível tirar melhor proveito de cada método.

Ao seguir essa combinação de metodologias, espera-se obter uma compreensão holística dos desafios éticos e responsabilidades associadas à utilização da inteligência artificial em engenharia, permitindo, assim, explorar as perspectivas dos envolvidos e descrever as características da expansão da IA.

1.3 Estrutura

O trabalho está dividido em 4 capítulos. O primeiro capítulo visa contextualizar o estudo com a apresentação e introdução da pesquisa, motivadores e justificativa para realização do trabalho. No segundo capítulo busca-se apontar os rumos de desenvolvimento da Inteligência artificial, bem como estabelecer seu conceito formal. No capítulo 3 pretende-se apresentar examinar e discutir as interações entre ética, inteligência artificial e Engenharia. Já no quarto capítulo são apresentadas as considerações finais de acordo com o objetivo traçado, discussão do trabalho e sugestões para futuras pesquisas.

2 O que há de real na Inteligência Artificial?

Para explorar os rumos da inteligência artificial na modernidade é necessário, antes de tudo, estabelecer um ponto de consenso onde o conceito de inteligência artificial seja delimitado e esclarecido para os propósitos que serão abordados nesta pesquisa. Isso se justifica, pois conforme firmado em estudos da filosofia, a delimitação conceitual pode influenciar a percepção da realidade e, com isso, torna-se possível elucidar questões relacionadas ao comportamento humano.

2.1 O que há de artificial na IA?

O conceito de inteligência, por si só, está enraizado na experiência e nos processos existentes em seres humanos, dada a capacidade de compreender, aprender, resolver problemas e outras habilidades cognitivas. De acordo com uma das definições mais influentes sobre inteligência, proposta por [Gardner \(1995\)](#), esta não pode ser reduzida a uma única habilidade geral, mas deve ser compreendida em termos diversos, relacionados a habilidades específicas.

No contexto da Inteligência Artificial (IA), entretanto, não se ancora necessariamente o termo na inteligência geral, uma vez que comumente se refere a aspectos da cognição humana. Portanto, é importante destacar que o conceito de IA, tal como elaborado nesta pesquisa, é baseado no teste proposto por Alan Turing em 1950 e citado no livro de [Russel \(2013\)](#) para definir, de maneira operacional, o conceito de inteligência, e em relação à engenharia, considerando-a como a busca pelo melhor programa de agente inteligente em uma dada arquitetura digital.

O teste de Turing busca verificar se uma máquina pode ser considerada inteligente ao exibir inteligência de forma suficientemente convincente em uma interface de texto, de modo que um questionador humano não consiga distinguir, em pelo menos 30 por cento do tempo de interação, se está interagindo com um participante humano ou uma máquina. Assim, o modelo de teste mais simples mede a capacidade do computador de: processamento de linguagem natural, representação de conhecimento, raciocínio automatizado e aprendizado de máquina.

Para fins de pesquisa, ao utilizar os conceitos de agente aprendido no curso, é coerente visualizar a IA como um agente racional ou inteligente, o mais próximo da racionalidade que pode ser atingido atualmente por meio de projetos. Assim, ao entrelaçar essa relação conceitual, pode-se afirmar que um agente racional tem o objetivo geral de maximizar a sua utilidade, considerando o seu histórico de percepção em determinada arquitetura. De outra forma, a IA como um campo de estudo se baseia na interseção entre a

psicologia cognitiva e as ciências computacionais. A proposta da IA é desenvolver sistemas artificiais capazes de apresentar ou simular determinadas características da inteligência.

Agente inteligente é um sistema capaz de perceber seu ambiente, processar informações, tomar decisões e agir para alcançar objetivos específicos. Sua estrutura geral básica está apresentada na Figura 2. Tais agentes podem variar em complexidade, desde agentes simples que operam em ambientes estáticos até sistemas altamente sofisticados que lidam com ambientes dinâmicos e incertos.

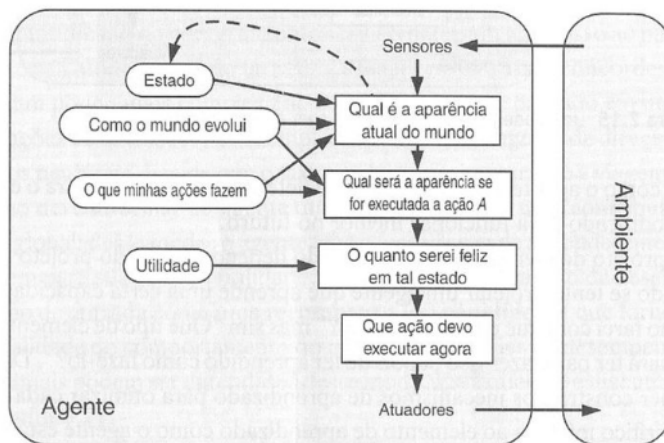


Figura 2 – Agente de utilidade baseado em modelos. Fonte: Russel (2013).

Assim, pode-se notar que os principais fatores para o projeto e funcionamento de um agente são dependentes da natureza do ambiente, como ações disponíveis para o agente (determinados por seu conjunto de atuadores), sua percepção (mediada pelos sensores) e comportamento específico, que é programado e estruturado em seu *softwares*. Todos esses aspectos devem estar envolvidos no projeto de engenharia de robôs reais ou ferramentas virtuais de IA, para que eles funcionem de maneira eficiente ao mesmo tempo em que se aplica a avaliação de seus impactos para futuros cursos de ação, principalmente quando se trata da ética.

É importante observar a narrativa da IA em várias formas de mídia, como o livro “Eu, Robô” de Asimov (2015), publicado pela primeira vez em 1939, uma obra de futuro distópico em que robôs estão totalmente integrados à sociedade. Vale citar também crítica ao capitalismo por meio do retrato de uma sociedade que pode ir a um parque temático para jogar com robôs “hiper-realistas” que, ao longo da série, adquirem uma espécie de auto-consciência, ou uma “consciência artificial”, além de outras produções cinematográficas que frequentemente retratam a IA como uma ameaça iminente à humanidade.

Essas representações são intrigantes e geram reflexões significativas sobre o futuro da tecnologia. No entanto, é crucial lembrar que a construção de sistemas de IA responsáveis, que atuem em conformidade com princípios éticos e considerem os valores humanos, é o foco principal. Embora se possam extrair lições dessas narrativas, é essencial não permitir que

preocupações exacerbadas com cenários distópicos desviem a atenção das oportunidades e dos desafios reais enfrentados na atualidade.

No entanto, existem concepções e objeções variadas quanto à viabilidade da IA na realidade. Essa divergência de opiniões baseia-se, em grande parte, na definição atribuída à IA. Dessa maneira, duas hipóteses famosas são consideradas: a primeira sugere que a máquina pode agir de maneira inteligente ao apresentar resultados e ações esperados de um sistema inteligente (tarefas específicas e limitadas), denominada de IA fraca ou estreita e a segunda sugere que as máquinas, além de exibirem comportamento, também são capazes de tomar decisões conscientes (replicando todas as habilidades intelectuais humanas), sendo denominada de IA forte ou completa.

Algumas objeções à possibilidade de existência de máquinas inteligentes serão abordadas em momentos oportunos nesta pesquisa, visto que já existem aplicações de IA no cotidiano, tais como os chamados “assistentes virtuais” (e.g. *Alexa* da Amazon e *Siri* da Apple), além sistemas de recomendação, carros autônomos e auxílio no diagnóstico médico por meio da análise de imagens. Essas aplicações ilustram a interseção vital entre a engenharia e a IA, evidenciando sua relevância e influência em diversas esferas da sociedade contemporânea.

2.2 Responsabilidade na utilização da inteligência de máquina

À medida que a IA se torna uma presença cada vez mais abrangente e influente em várias esferas da sociedade, surge uma demanda crítica por abordagens éticas e responsáveis para lidar com os desafios decorrentes dessa expansão. A visão sobre esses aspectos pode ser orientada pelo marco da “Declaração Universal dos Direitos Humanos” das Nações Unidas, adotada em 1948, que estabelece valores universais que afirmam a dignidade intrínseca e os direitos inalienáveis de todos os seres humanos, como fundamento da liberdade, da justiça e paz no mundo ¹ A declaração constitui um marco importante ao destacar a igualdade, a liberdade e a justiça como direitos fundamentais e, por extensão, pode-se entender que sustenta a necessidade de que o desenvolvimento e uso da IA respeitem esses princípios essenciais.

Ademais, a ética e a filosofia desempenham um papel crucial na definição da responsabilidade na utilização da IA. A ética, baseada na reflexão sobre o que é moralmente correto e justo, busca orientar as ações humanas de acordo com princípios que promovem o bem-estar e a dignidade de todos os indivíduos. Por sua vez, a filosofia oferece um ponto de partida do estudo intelectual necessário para analisar as complexas implicações éticas da IA, examinando seus impactos na sociedade, na tomada de decisões e nas interações humanas.

¹ [Unicef \(2023\)](#).

Um dos principais dilemas associados à IA é a dificuldade em identificar responsabilidades claras em casos de erros ou consequências indesejadas decorrentes das ações de sistemas automatizados. Nota-se que as atuais aplicações de IA muitas vezes não possuem uma identificação explícita de “pais” ou criadores, tornando complexa a atribuição de responsabilidade jurídica. Esse cenário é especialmente relevante quando falhas ocorrem, seja no diagnóstico médico, na produção industrial ou em interações virtuais, levantando questionamentos sobre quem deve ser responsabilizado por esses resultados.

Essa dificuldade é chamada de **opacidade** e, em contexto de sistemas de inteligência computacional, refere-se à dificuldade de compreender completamente como um sistema toma decisões ou chega a determinadas conclusões. É a falta de transparência sobre o processo interno do algoritmo, tornando difícil para os usuários ou observadores entenderem o raciocínio por trás das ações da IA. A opacidade pode ser um desafio significativo em termos de explicabilidade e responsabilização dos sistemas de inteligência artificial.

Segundo o estudo sobre falhas e vieses algorítmicos realizado por [Alves e Andrade \(2021\)](#) “... a opacidade pode encobrir e potencializar falhas e vieses em algoritmos”. Essa opacidade pode ser comparada a uma “caixa-preta” algorítmica que necessita ser iluminada para possibilitar aos desenvolvedores a identificação e correção de possíveis falhas e vieses.

A coleta massiva de informações por parte das empresas e organizações também levanta preocupações quanto à privacidade e à proteção dos dados pessoais dos indivíduos. As práticas de coleta, armazenamento e compartilhamento de dados, muitas vezes realizadas sem o pleno conhecimento dos usuários, suscitam questões sobre a transparência e o consentimento informado. Além disso, o comércio eletrônico e as plataformas digitais têm revolucionado a forma como produtos e serviços são comercializados, proporcionando conveniência e acesso global, mas também gerando debates sobre a equidade nas transações, concorrência justa e os impactos socioeconômicos nos mercados tradicionais. Isso torna indispensável o fornecimento de um quadro ético para guiar as práticas comerciais, garantindo a proteção dos direitos individuais e a promoção de relações comerciais justas e benéficas para todas as partes envolvidas

Nesse sentido, conforme citado por [Nilsson \(1998\)](#) em seu livro sobre IA e a elaboração da sua arquitetura e *softwares*, é essencial estabelecer normas e diretrizes claras para garantir a utilização responsável e ética da IA em projetos e processos de engenharia, assegurando que benefícios e riscos sejam adequadamente equilibrados, já que essa automação também pode resultar em dilemas éticos, uma vez que decisões cruciais podem ser delegadas a algoritmos que carecem de compreensão moral e discernimento humano.

Além das questões de responsabilidade direta, a expansão da IA também traz à tona desafios éticos e sociais que merecem atenção cuidadosa. A aplicação da IA pode resultar em situações complexas, como discriminação algorítmica, viés de confirmação,

disseminação de desinformação e até mesmo ações prejudiciais, como o *cyberbullying*, demandando a implementação de técnicas para mitigar tais vieses e garantir a equidade nas decisões tomadas pelos sistemas de IA. Ademais, a falta de discernimento moral por parte das máquinas e a capacidade de reproduzir preconceitos presentes nos dados de treinamento podem desencadear impactos negativos e imprevisíveis na sociedade.

Dessa forma, não é exagero reafirmar a importância de estabelecer diretrizes éticas sólidas e regulamentações adequadas para orientar o desenvolvimento, a implementação e o uso da inteligência artificial. A definição de padrões éticos para a programação de sistemas de IA, bem como a criação de mecanismos de supervisão e responsabilização são passos cruciais para garantir que os benefícios da IA sejam maximizados enquanto se minimizam os riscos e impactos negativos.

Portanto, os engenheiros desempenham um papel fundamental na construção de sistemas de IA responsáveis e éticos, ao considerarem não apenas os aspectos técnicos, mas também os impactos sociais, culturais e éticos de suas criações. Essas medidas não apenas protegem os usuários e a sociedade como um todo, mas também contribuem para a construção de um ambiente em que a IA seja uma força positiva e transformadora, alinhada com princípios éticos sólidos.

2.3 Uma Filosofia para a Informação

Nesta seção, será explorada parte da teoria e os estudos do filósofo Luciano Floridi, professor de Filosofia e Ética da Informação na Universidade de Oxford, que emergiu como uma figura de destaque no campo da ética e filosofia da informação, oferecendo uma perspectiva única sobre as implicações éticas e conceituais da inteligência artificial (IA) e sistemas autônomos. A obra de Floridi oferece uma perspectiva abrangente e profunda sobre os desafios éticos e conceituais que surgem com o avanço dessas tecnologias e volta nosso olhar para questões que são mais pupáveis na atualidade.

Suas contribuições têm um impacto significativo e transdisciplinar, que abrange diversos campos do conhecimento humano, incluindo ética da informação, ética da computação e filosofia da IA. Em particular, suas teorias sobre a Ontologia da Informação e a relação entre informação, significado e contexto são essenciais para compreender como os sistemas autônomos e inteligentes interagem com o ambiente e como essas interações levantam questões éticas.

Ao falar de Floridi, é preciso realizar uma breve introdução a Filosofia da Informação, um campo de pesquisa em construção que, de acordo com [De Gómez \(2013\)](#), para muitos deveria ser perguntado primeiro se a informação por si mesma, poderia ser um objeto de problematização filosófica. Justamente neste ponto Floridi defende a dignidade filosófica da informação em si, o que significa que os problemas estudados por ele muitas vezes não

serão “sobre” essa filosofia mas “desse” campo de estudo com uma perspectiva das ciências da computação, que muitas vezes revela a apropriação seletiva da filosofia analítica.

Com isso, seu trabalho tem como corpo filosófico a Teoria da Ontologia da Informação, que explora a natureza da informação, seu significado e como ela está integrada ao mundo físico e digital. A ontologia é um campo interdisciplinar que se concentra na natureza e na organização da informação, em Floridi (2010, p.11) ele a define como um termo para se referir a uma forma que transforma fundamentalmente a natureza típica de um sistema. Ela investiga as categorias fundamentais de entidades e relações que compõem o universo informacional, buscando compreender a estrutura e o significado subjacentes aos dados e às informações. Ela oferece uma estrutura conceitual para explorar como os sistemas autônomos e inteligentes interagem com o ambiente e como essa interação levanta questões éticas.

Um conceito central na teoria da Ontologia da Informação é o *Standard Definition of Semantic Information (SDI)*. O SDI consiste em uma definição padrão de informação semântica, que se refere à informação que tem significado ou sentido, ou seja, que possui conteúdo e contexto que pode ser compreendido e interpretado por um agente cognitivo, como um ser humano ou um sistema inteligente. Para exemplificar essa definição, um dado com significado em si pode ir de um semáforo de trânsito, que dependendo da cor define se o condutor deve seguir ou parar; a temperatura, que possui um significado e contexto que podem ser compreendidos para tomar decisões, como escolher roupas apropriadas e as cores de frutas, fornecem informações semânticas sobre o estágio de maturação e sabor da fruta. Em outras palavras, a informação semântica carrega significado e é capaz de contribuir para a compreensão e a tomada de decisões.

Dessa forma, se torna mais fácil compreender a própria definição de Floridi sobre a filosofia da informação:

“... A filosofia da informação é o campo filosófico concernente com (a) a investigação crítica da natureza conceitual e os princípios da informação, incluindo suas dinâmicas, utilização e ciências, e (b) a elaboração e aplicação de metodologias teóricas, informacionais e computacionais, a problemas filosóficos.”²

Seguindo tal abordagem, a informação adquire significado por meio do contexto e das relações, e como isso é fundamental para a compreensão do mundo. Assim, ela não se trata apenas de um conjunto de *bits* ou sinais, o que seriam elementos brutos ou os dados para Floridi, mas uma entidade que carrega consigo um significado relevante para um observador ou agente.

Além disso, a abordagem de Floridi destaca a importância do contexto na interpretação da informação, o que gera o conhecimento, em sua acepção. O mesmo conjunto de

² Floridi (2020, p.14).

dados pode ter significados diferentes em contextos distintos. Ele propõe que a informação semântica é caracterizada por sua relevância e utilidade para um observador, levando em consideração o contexto no qual a informação é recebida e interpretada. Essa abordagem traz consigo implicações da filosofia da informação e outros campos que são relevantes nesta pesquisa: inteligência artificial e ética, pois são espaços onde a compreensão do significado do dado é crucial para tomada de decisões.

Com essa base, já se pode dar continuidade a dois princípios fundamentais da teoria que se mostram relevantes neste trabalho: ética da informação e o ciberespaço como ambiente de vida. Do primeiro, tem-se a estreita relação da ética da informação com a ética da computação, que seria responsável por tratar das práticas, comportamentos e responsabilidades no uso de sistemas computacionais. Com respeito ao segundo, considera-se o surgimento de novos desafios éticos na era digital, incluindo questões relacionadas à disseminação, privacidade, segurança, propriedade intelectual e acesso à informação.

A ética da informação, segundo o pensador, não se concentra apenas nas ações humanas no mundo digital, mas também na própria natureza da informação e na maneira como ela está inserida na estrutura da realidade. Isso inclui considerar como a informação é produzida, disseminada, acessada e utilizada, bem como as implicações éticas de sua manipulação. Nesse contexto, ele introduz o conceito de “infomoralidade”, refere às ações moralmente relevantes que ocorrem no ambiente digital que devem ser avaliadas, assim como as ações no mundo físico não-digital.

Floridi também destaca a importância do ciberespaço como um ambiente de vida no seu livro “[The fourth revolution: how the infosphere is reshaping human reality \(2014\)](#)”. Na referida obra, o autor descreve como o ciberespaço não deve ser considerado simplesmente como um espaço isolado, mas sim como um ecossistema intrinsecamente conectado, onde as fronteiras entre o mundo virtual e o mundo físico se tornam cada vez mais tênues. Argumenta ainda que o ciberespaço está totalmente entrelaçado com as atividades diárias das pessoas, incluindo o uso de dispositivos móveis para comunicação, o acesso a informações e serviços online, compras pela internet, interações em redes sociais e muito mais.

Essa visão do ciberespaço como parte integrante da vida contemporânea é crucial para a compreensão das complexidades éticas que surgem com o avanço das tecnologias digitais e influencia a forma como se entende a interdependência crescente entre o mundo virtual e o mundo físico e desafia a abordar o desenvolvimento e a aplicação da tecnologia de forma responsável.

Como forma de estimular debates sobre essas questões e abordar implicações da era da informação na sociedade contemporânea, foi publicado o livro “[The Onlife Manifesto \(2015\)](#)” como um projeto colaborativo envolvendo diversos acadêmicos, incluindo o próprio Floridi. Conforme o estudo, a ideia central é enfatizar a necessidade de uma abordagem

holística, considerando o sistema como um todo integrado e interdependente, reconhecendo que mudanças em uma área podem ter impactos em outras áreas; e interdisciplinar para entender e enfrentar os desafios e oportunidades dessa nova era.

Nesta obra destaca-se o conceito de “... ciberespaço como ambiente de vida”, proposto pelo autor, em que se verifica a influência de interações, decisões e experiências em grande escala: essa perspectiva destaca a maneira pela qual a tecnologia digital permeia a existência contemporânea, sendo ela nas questões inerentes à comunicação, percepções de mundo e até mesmo relações interpessoais.

De modo geral, ao fazer um recorte de classe social e ambiente, é raro ver um universitário que não esteja conectado a internet ou recortando ainda mais, a um dispositivo móvel. A variedade de dispositivos eletrônicos e preços torna-os acessíveis a diversos estratos sociais. Uma ilustração concreta dessa interconexão entre o real e virtual pode ser observada na reconfiguração do mercado e do varejo, que passou por mudanças significativas para atender às demandas comerciais dos consumidores.

O impacto do comércio eletrônico ou *e-commerce* alterou substancialmente a dinâmica desses setores, ainda que não tenha ocorrido de maneira uniforme em todas as regiões geográficas, como destacou o estudo de [Galinari Rangel \(2015\)](#), conduzido pelo BNDES. Ao considerar o ciberespaço como um ambiente de vida, é possível reconhecer não apenas oportunidades e benefícios, mas também as complexidades éticas e sociais que surgem desse entrelaçamento cada vez mais profundo entre a esfera digital e a vida real.

Os autores especificam ainda quatro aspectos da vida humana que já foram profundamente impactados pela transição digital:³

1. Confundindo a distinção entre realidade e virtualidade;
2. Apagando as distinções entre humano, máquina e natureza;
3. Indo da escassez para abundância, no que diz respeito à informação;
4. Passando da primazia das entidades sobre as interações para a primazia das interações sobre as entidades.

Floridi argumenta que, se não se considerarem cuidadosamente essas mudanças e desafios, pode-se oscilar entre a desconfiança na tecnologia e uma fé cega nela, ambas insuficientes para guiar uma vida pública significativa. Por isso, esses quatro aspectos servirão como pontos de referência fundamentais para orientar esta pesquisa.

Portanto, as ideias e teorias de Luciano Floridi fornecem um suporte valioso para compreender as implicações éticas e filosóficas da inteligência artificial e dos sistemas

³ Tradução própria de trecho disponível em [Floridi \(2015\)](#).

autônomos. Sua abordagem interdisciplinar, enraizada na filosofia da informação e teoria da Ontologia da Informação fornecem uma estrutura conceitual para abordar questões éticas e filosóficas na criação, implementação e uso de sistemas inteligentes, o que enriquece o debate sobre como lidar com as complexidades éticas e conceituais que surgem com o avanço dessas tecnologias, orientando a busca por um desenvolvimento e aplicação responsáveis da IA na sociedade contemporânea.

2.4 IA na Engenharia

Os sistemas chamados inteligentes têm se estabelecido como um campo interdisciplinar que abrange diversas áreas do conhecimento, e a engenharia desempenha um papel fundamental neste cenário ao desenvolver estruturas, algoritmos e sistemas que viabilizam a construção e aplicação da IA em uma ampla gama de setores. Além disso, é possível observar que essa correlação aumenta se for considerado que a IA não é apenas um campo teórico, mas também prático, com aplicações que permeiam diversas esferas da sociedade, incluindo automação industrial, saúde, transporte e educação.

Portanto, pode-se inferir que a engenharia entra em cena como o catalisador que transforma os conceitos teóricos da IA em realidade tangível. É por meio da engenharia que os algoritmos são desenvolvidos, os modelos são construídos e os sistemas são implementados. No livro “[Inteligência Artificial \(2013\)](#)”, são exploradas algumas aplicações práticas da IA que ilustram como a engenharia se entrelaça neste processo. A obra ressalta que a implementação da IA envolve a concepção de algoritmos inteligentes, a engenharia de sistemas de aprendizado de máquina e a criação de ambientes de interação homem-máquina. Esses aspectos práticos encontram seu alicerce na engenharia, onde as teorias são traduzidas em códigos, dispositivos e sistemas reais.

Neste contexto, a Inteligência Artificial juntamente com a engenharia emerge como um fator de destaque na reconfiguração da indústria, desempenhando um papel fundamental na era da Revolução Industrial 4.0.⁴ A Figura 3 ilustra de maneira breve a evolução da Indústria até a chamada “Indústria 4.0”, e a pesquisa de [Sakurai e Zuchi \(2018\)](#) delinea algumas facetas em destaque dessa transformação, entre elas a integração com a Internet das Coisas, o fortalecimento da segurança cibernética, a análise dos *big data* e o avanço da robótica.

Além do mais, é importante destacar que o termo “Quinta Revolução Industrial”, embora ocasionalmente discutido, ainda não possui uma definição completa ou é ampla-

⁴ A engenharia de IA abrange uma série de disciplinas, desde a programação de *softwares* até o design de hardware, visando criar sistemas inteligentes capazes de perceber, interpretar e interagir com seu ambiente, assim como tomar decisões com base em dados e informações. Muitos desses processos podem ser encontrados e descritos no livro [Inteligência Artificial \(2013\)](#), que se constitui como uma excelente referência para estudos da área.

mente reconhecido como uma etapa distintiva na evolução industrial. Até o momento, a terminologia mais aceita e utilizada para descrever o atual avanço tecnológico e industrial permanece sendo a Quarta Revolução Industrial.



Figura 3 – Evolução industrial.

Do mesmo modo, a universidade, como ambiente de pesquisa e ensino, desempenha um pilar da integração da engenharia com a IA. Já que ela é um espaço onde teoria e prática devem se entrelaçar, permitindo que estudantes e pesquisadores explorem e experimentem os princípios da IA por meio de projetos de pesquisa teóricos e práticos, como o presente estudo. Não obstante, a universidade é um terreno fértil para discussões éticas e filosóficas sobre o uso da IA, proporcionando um ambiente propício para a reflexão crítica e a busca por diretrizes mais sólidas e responsáveis no desenvolvimento e aplicação da IA, como destaca o filósofo em seu livro “[AI and its new winter: from myths to realities \(2020\)](#)”.

A engenharia ocupa um papel central na aplicação prática da IA, ao traduzir conceitos teóricos em soluções concretas, impulsionando a inovação e permitindo que os benefícios da IA se manifestem de maneira palpável na sociedade. Por meio de uma colaboração entre a engenharia e a IA, o potencial transformador dessas tecnologias é explorado e materializado em benefício de diversas áreas e setores.

Portanto, neste casamento entre tecnologia e engenharia sobressai o potencial da IA em aprimorar processos industriais, ao mesmo tempo em que enfatiza a necessidade de considerar as implicações éticas de sua aplicação, garantindo a preservação da dignidade humana em todas as fases do desenvolvimento e implementação da IA.

Controle & Automação

No ramo específico da engenharia voltado aos sistemas de controle e automação, é inquestionável a importância da IA, sobretudo na otimização de processos industriais. A Engenharia de Automação Industrial é um campo em constante evolução, e a IA desempenha um papel fundamental nesse avanço ao permitir que máquinas e sistemas tomem decisões autônomas, “aprendam” com experiências passadas e se adaptem a novos desafios em tempo real.

O Aprendizado de máquina ou *Machine Learning* (ML) é um subcampo da inteligência artificial que permite que sistemas automatizados aprendam com dados ou experiências passadas, aprimorando seu desempenho ao longo do tempo. Essa abordagem tem a capacidade de melhorar o desempenho dos algoritmos e resultados das aplicações em diversos campos.

De modo geral, os modelos de aprendizado de máquina podem ser classificados de várias maneiras, e seu funcionamento tem como inspiração do aprendizado humano, que pode ocorrer por meio de inferências indutivas, como Arquimedes com o princípio da alavanca e a primeira lei da hidrostática.

Nesse contexto, o ML tem o aprendizado indutivo como maneira de se classificar seu sistema de aprendizado. Pode-se dividir esse sistema em três categorias principais, com base na quantidade de supervisão durante o treinamento do modelo: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço, conforme ilustrado na Figura 4.

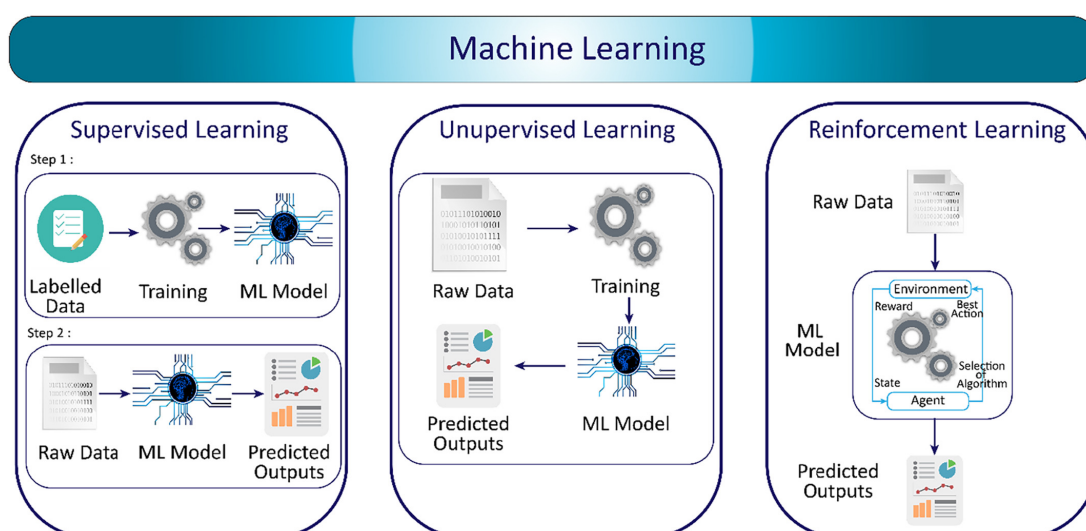


Figura 4 – Agente de utilidade baseado em modelos. Fonte: [Baduge et al. \(2022\)](#).

Seguindo a representação da Figura 4 e as diretrizes de [Monard e Baranauskas \(2003\)](#) no aprendizado supervisionado, fornece-se ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido.

No aprendizado não supervisionado, um algoritmo de inteligência artificial analisa um conjunto de dados sem informações prévias sobre as categorias ou grupos existentes. Ele busca identificar padrões ou relações entre os elementos, agrupando-os em clusters ou conjuntos com base em características compartilhadas. Após essa etapa de agrupamento, requer-se uma análise adicional para compreender o significado no contexto em questão.

A aprendizagem por reforço se configura como um método de ML que capacita o

modelo por meio de um sistema de recompensas e penalidades. Esse sistema de aprendizagem “observa” o ambiente e executa ações visando aprimorar sua estratégia para alcançar o desempenho ideal.

Por exemplo, no caso da IA de respostas a perguntas como o *ChatGPT* – que fazem parte dos chamados modelos de linguagem generativa – a tecnologia de aprendizado de máquina permite que o sistema aprimore continuamente sua capacidade de entender e responder às perguntas dos usuários, tornando-o cada vez mais preciso e útil.

Em cenários de automação industrial, a capacidade de aprendizado de máquina em tempo real significa que os sistemas poderiam se adaptar instantaneamente às mudanças nas condições da fábrica, otimizar processos em tempo real e tomar decisões ágeis para garantir a eficiência, a segurança e a qualidade da produção.

Essa tecnologia poderia ser aplicada em várias etapas da produção industrial, desde o monitoramento e controle de máquinas até a otimização de cadeias de suprimentos e logística. Os sistemas de IA podem analisar grandes volumes de dados em tempo real, identificar padrões, prever falhas e tomar decisões com base em algoritmos avançados tornando as máquinas e processos mais eficientes e práticas mais sustentáveis.

Um exemplo notável é a utilização de robôs autônomos em linhas de produção com o uso da inteligência artificial como na fábrica da Tesla, a fabricante de veículos elétricos. Conforme o [Tesla Factory Tour with Elon Musk! \(2023\)](#) (*tour* pela fábrica com Elon Musk), a Tesla utiliza robôs com visão computacional e algoritmos de aprendizado de máquina para automatizar grande parte do processo de montagem de veículos.

Em suas operações, os robôs são capazes de realizar tarefas complexas, como soldagem e montagem de peças, com precisão e eficiência. A empresa também aproveita dados em tempo real de sensores em seus veículos em operação para melhorar constantemente o desempenho e a segurança.

No entanto, essa transformação também traz consigo desafios éticos e responsabilidades significativas que merecem atenção cuidadosa neste campo. À medida que máquinas e sistemas autônomos desempenham um papel mais central nas operações industriais, é necessário que os princípios éticos sejam incorporados em sua concepção e uso. Uma vez que as decisões tomadas por esses sistemas podem afetar a qualidade dos produtos, a segurança dos trabalhadores e a sustentabilidade dos processos.

Assim, a pesquisa pretende contribuir para o avanço da automação industrial de forma ética e responsável, assegurando que a IA seja uma aliada na busca por processos mais eficientes, sustentáveis e seguros, enquanto respeita os valores fundamentais da dignidade humana, por meio de uma abordagem crítica e embasada nas obras de pesquisadores como Luciano Floridi e outros especialistas em ética da IA.

3 O verão da Inteligência Artificial

Este capítulo visa examinar e discutir as interações entre ética, Inteligência Artificial (IA) e Engenharia. Essa análise se fundamenta no alinhamento prévio estabelecido sobre o papel de cada um desses temas para a pesquisa, considerando também o aporte teórico escolhido.

Conforme a ideia desenvolvida por John McCarthy,¹ um dos pioneiros da IA, o “Verão da Inteligência Artificial” pode ser entendido como um período caracterizado pelo entusiasmo, desenvolvimento acelerado e significativos investimentos em novas soluções baseadas em aprendizado de máquina.

3.1 O desafio da Automação onipresente

Uma análise inicial revela uma divulgação crescente de ferramentas acessíveis ao público em geral, como o *ChatGPT*, a assistente de voz *Alexa* e outros *chatbots*. Tais ferramentas compartilham a característica de serem sistemas baseados em IA que utilizam processamento de linguagem natural para interagir com os usuários e se integram a uma variedade de dispositivos e plataformas.

O modelo *ChatGPT* desenvolvido pela *OpenAI*,² por exemplo, constitui uma inteligência artificial baseada em linguagem que emprega uma arquitetura de aprendizado profundo para gerar respostas e interagir naturalmente com os usuários. A *Alexa*, por sua vez, conforme definido pela própria empresa criadora, Amazon, é uma inteligência artificial controlada por voz, “residindo na nuvem” e capaz de prestar assistência em ambientes conectados à Internet.³

Apesar dessas semelhanças, existem divergências no escopo de uso e capacidades específicas de cada tecnologia que são evidentes. A *Alexa* concentra-se em tarefas práticas de controle de dispositivos no cotidiano, enquanto o *ChatGPT* está direcionado para interações de linguagem natural e respostas a perguntas, lembrando, em termos de usabilidade, o reconhecido do motor de busca do Google. Uma análise das tendências de busca no *Google Trends* reforça essa distinção,⁴ evidenciando picos de interesse por “*Alexa*” próximos a lançamentos, enquanto que o *ChatGPT* ultrapassou, em abril de 2023, quase quatro vezes o maior índice de busca da *Alexa*.

Nesse contexto, nota-se que, embora a *Alexa* tenha desempenhado um papel

¹ Ver em Floridi (2020).

² Ver em OpenAI (2023).

³ Amazon (2023).

⁴ GOOGLE... (2023).

significativo no desenvolvimento e evolução da IA popular para uso próprio, principalmente no que diz respeito à acessibilidade e interação natural em aspectos da vida cotidiana, foi o *ChatGPT* que efetivamente marcou o fim do possível “inverno da IA” para além dos meios industriais e acadêmicos. Sua notável capacidade de gerar respostas em linguagem natural e interagir de maneira fluida e realista com os usuários desempenha um papel crucial nessa nova fase da evolução da inteligência artificial.

Assim, a introdução da tecnologia no cotidiano hoje pode ser vista com um processo natural, como observado por Floridi (2014), que descreve o ciberespaço como “... um ambiente de vida”, isto é, aquele lugar criado pela interconexão global de redes de computadores e dispositivos eletrônicos, e seus respectivos meios de funcionamento, digital ou virtual. Mas essa aparente neutralidade da tecnologia esconde uma realidade mais complexa, como asseveram Novaes e Dagnino (2004). Em seu trabalho é ancorada na visão de Andrew Feenberg,⁵ que apresenta o avanço tecnológico como algo intimamente ligado às relações sociais, por vezes refletindo a influência da produção capitalista e negligenciando a participação democrática de diversos grupos na definição da tecnologia.

Ainda no artigo, os autores sintetizam o pensamento de Feenberg destacando que a tecnologia não é neutra, pois incorpora valores da sociedade industrial e, frequentemente, reflete os interesses das elites que têm influência sobre seu desenvolvimento. Em seu livro “*The ethics of artificial intelligence* (2014)”⁶ Nick Bostrom, professor na Universidade de Oxford e fundador do Instituto Future of Humanity, e outros autores, exploram a ideia de que o avanço da IA nem sempre mostra novos desafios éticos, pois quando um algoritmo de IA é utilizado em atividades cognitivas com dimensões sociais que previamente eram realizadas por humanos, ele herdará os requisitos e valores éticos de seus criadores, podendo ocasionar novos desafios de programação e desafios éticos como um espelho dos valores de determinadas coletividades.

Em tais cenários a tecnologia pode ser vista não apenas como um produto técnico, mas também um veículo para a dominação cultural, controle social e concentração de poder. Portanto, a racionalidade técnica está intrinsecamente ligada à racionalidade política. Nesse contexto, emerge e se manifesta o conceito de “Fetiche da Tecnologia”, em que a tecnologia em si é vista como um avanço inquestionável e sempre benéfico. No entanto, como apontado no Novaes e Dagnino (2004), muitas vezes negligenciamos as relações de dominação que continuam a moldar a sociedade e a tecnologia.

Destacando que as decisões tecnológicas não são meramente baseadas em critérios técnicos, mas frequentemente incorporam os valores do capitalismo e fortalecem a acumulação de capital, de onde surgem preocupações entre os trabalhadores sobre a substituição

⁵ Andrew Feenberg (nascido em 1943) é um filósofo da tecnologia americano. Ele ocupa a Cátedra de Filosofia da Tecnologia na Escola de Comunicação da Universidade Simon Fraser, em Vancouver, Canadá.

⁶ Bostrom e Yudkowsky (2014, p. 317).

da mão de obra humana pelo uso de máquinas e o desemprego estrutural. Tal desequilíbrio entre a demanda e a oferta de trabalho em determinados setores ou regiões de uma economia é persistente.

Ao discutir o desemprego em meio à automação impulsionada pela IA, é importante analisar a visão de Luciano Floridi, renomado filósofo da tecnologia. [Floridi \(2014\)](#) destaca que a automação promete maior eficiência e produtividade, mas também levanta questões fundamentais sobre o impacto na força de trabalho e nas relações sociais. Argumenta ainda que a sociedade está diante do desafio da automação, e que a implementação dessas tecnologias exige uma revisão e atualização constante das medidas de proteção.⁷

Neste aspecto, as perspectivas de Feenberg e Floridi são cruciais para que se compreenda o contexto do desemprego frente à crescente integração da IA nas operações e processos de produção e o aumento na utilização cotidiana pelos cidadãos.⁸ Elas constituem um alerta para a necessidade de uma abordagem crítica e ética na implementação dessas tecnologias, visando não apenas à eficiência econômica, mas também ao bem-estar e à justiça social.

Desta maneira, a tecnologia também representa um contexto de construção social, um campo de batalha historicamente determinado, e a trajetória da inovação científica e tecnológica pode ser redirecionada pelo interesse na mudança social, interferindo tanto na divisão do trabalho e nos avanços éticos, quanto no processo decisório da política científica e tecnológica.

Nesse ponto, torna-se evidente que os sistemas de automação encontram-se intrinsecamente entrelaçados com a história e aplicação da Inteligência Artificial. Isto é, de certa forma, a história da Automação é também a história da Inteligência Artificial. Preocupações emergentes concentram-se nas implicações sociais e éticas desse progresso tecnológico. A sociedade encontra-se diante do desafio da automação, a qual promete não apenas maior eficiência e produtividade, mas também levanta questões fundamentais sobre o impacto na força de trabalho e nas relações sociais. Este cenário é claramente ilustrado na pesquisa sobre a legislação brasileira nesse âmbito realizada por [Martinez e Maltez \(2019\)](#).⁹

3.2 IA no ambiente de trabalho

A crescente incorporação de sistemas de IA nas operações e processos de produção pode desempenhar tarefas mais complexas e perigosas, anteriormente realizadas por

⁷ [Floridi](#).

⁸ Ver em [Novaes e Dagnino \(2004\)](#).

⁹ No entanto, é crucial salientar que as transformações trazidas pelo avanço da Inteligência Artificial no ambiente de trabalho demandam uma constante revisão e atualização dessas medidas de proteção, e esse aspecto será abordado ao longo do texto.

trabalhadores humanos. Isto poderá ocasionar, possivelmente, uma redução de riscos no ambiente de trabalho e, ao mesmo tempo, aumentar a produtividade do meio industrial, ou mesmo reconfigurá-lo por completo. Entretanto, continua gerando preocupações sobre o futuro do emprego e a segurança no local de trabalho. Portanto, é incentivada a realização de pesquisas sobre a regulamentação das transformações da IA no ambiente de trabalho em um cenário em constante evolução.

No contexto abordado, é pertinente destacar que mesmo em obras metodológicas, como apontado por [Russel \(2013\)](#), que ao explorar o conceito de ética em computação, examina, dentre várias questões, a preocupação com a possível substituição de empregos por automação em sua integração com a área da Inteligência Artificial. Como ressaltado em relação à chamada “Quarta Revolução Industrial” ou “Indústria 4.0”, a economia vem se tornando cada vez mais dependente de sistemas de computação e de aprendizado de máquina em diversas esferas, como, por exemplo, na avaliação de concessão de crédito aos consumidores apresentada por [Lima et al. \(2009\)](#), em que sistemas de IA baseados em redes neurais desempenham um papel crucial.

Em contrapartida, [Russel \(2013\)](#) apresenta uma visão puramente otimista da automação por meio da tecnologia de IA, ao informar que ela tem gerado um saldo positivo na criação de empregos, resultando em ocupações mais complexas. Afirma ainda que ela proporciona às pessoas a oportunidade de aumentar sua remuneração ao utilizarem esses agentes inteligentes. Isso se dá tanto pela análise dos sistemas quanto pela criação de ambientes de trabalho mais seguros e confortáveis, onde a IA assume o controle de tarefas de alto risco e, assim, colabora para preservar a integridade física dos trabalhadores.

Para além das questões sociológicas e de desemprego decorrentes do avanço no uso da tecnologia e da IA nos processos de fabricação, destaca-se o uso dos sistemas de tomada de decisão automatizada, chamados de *Automated Decision-Making Systems*, ou simplesmente *ADMS*. Esses sistemas estão sendo utilizados em diversos setores e isso significa que muitas decisões que trazem grandes impactos no ambiente ou vida das pessoas que antes eram tomadas por seres humanos hoje acontecem por sistemas ADMS. Na área da medicina, por exemplo, encontram-se sistemas de detecção de doenças, como a retinopatia diabética em fotografias de fundo de olho,¹⁰ ou a classificação do nível de câncer de pele.¹¹ Da mesma forma em setores automobilísticos que envolvem a criação de veículos autônomos,¹² e também de produção e plantio onde se dá a identificação eletrônica para agricultura de precisão,¹³ entre outras áreas de aplicação.

De acordo com a pesquisa de [Mokander et al. \(2021\)](#), o uso do termo ADMS confere um significado mais preciso dentro de IA e se refere aos sistemas sociotécnicos que

¹⁰ [Gulshan et al. \(2016\)](#).

¹¹ [Esteva et al. \(2017\)](#).

¹² [Gerdes e Thornton \(2016\)](#).

¹³ [Sophocleous \(2021\)](#).

englobam um modelo de tomada de decisão, um algoritmo que traduz esse modelo em código computável, os dados utilizados por esse código e o ambiente completo que envolve sua utilização.

Esses sistemas estão sendo progressivamente adotados por grandes empresas, tanto no setor público quanto privado, com o intuito primário de aprimorar a eficiência, escalabilidade e agilidade na busca por soluções inovadoras.

Seus mecanismos utilizam árvores de decisões, que são modelos gráficos que representam escolhas e seus possíveis resultados, ou até mesmo redes neurais profundas, sistemas de aprendizado de máquina inspirados na estrutura do cérebro humano.

No livro de Russel (2013) pode-se encontrar uma descrição acadêmica das formas em que os mecanismos de IA podem ser projetados, para esta pesquisa se torna necessário o destaque das árvores de decisão que permite que o sistema avalie diferentes caminhos de ação com base em critérios predefinidos, e as redes neurais profundas, que são capazes de aprender com dados e identificar padrões complexos.

Para compreender como esses mecanismos de decisão funcionam, é fundamental analisar o processo subjacente. Em geral, eles recebem um conjunto de dados de entrada, que pode incluir informações contextuais, históricos ou indicadores relevantes. Em seguida, o sistema utiliza algoritmos e modelos matemáticos para processar esses dados e atribuir valor ou probabilidade a diferentes opções de decisão.

À medida que mais dados são processados e mais decisões são tomadas, esses sistemas podem aprender e aprimorar sua capacidade de tomar outras decisões precisas. No entanto, é importante destacar que a transparência e a ética desempenham um papel crítico nesse contexto, uma vez que decisões automatizadas podem afetar significativamente a vida das pessoas e as operações das organizações.

O número de dados que são disponibilizados atualmente chegam a números gigantesco se comparados com o passado, por isso Floridi (2014) destaca a sobreabundância de informações na era digital. Nesse contexto, a ética na IA mostra-se crucial, especialmente no que diz respeito à segurança e responsabilidade dos dados.

Hoje a disponibilidade de dados pessoais cresce consideravelmente por meio do uso generalizado de redes sociais, aplicativos de celular e outros serviços que armazenam informações pessoais e de acesso, muitas vezes utilizadas para direcionar anúncios ao público, sem que os usuários tenham total clareza sobre os limites desse compartilhamento e alcance dos dados.

No Brasil, mesmo com o Marco Civil da Internet, Lei nº 12.965/2014, que busca proteger os direitos da personalidade nas redes sociais virtuais e promover pesquisas sobre a privacidade e responsabilidade civil.¹⁴ À medida que a disponibilidade de dados cresce

¹⁴ Tefé e Moraes (2017).

exponencialmente, a responsabilidade na coleta, armazenamento e uso de informações torna-se ainda mais crucial e ao mesmo tempo complexa. Esse processo exige uma atenção cuidadosa para assegurar a segurança e a proteção dos direitos individuais, bem como alinhando-se com a Lei Geral de Proteção de Dados (LGPD) do Brasil,¹⁵ que estabelece diretrizes específicas para o tratamento de dados pessoais no país.

Para ilustrar o uso desses dados tem-se, por exemplo, que a decisão de aprovar ou negar uma hipoteca com base em critérios algorítmicos pode levantar questões sobre discriminação e justiça. Da mesma forma, os algoritmos em veículos autônomos podem enfrentar dilemas éticos ao decidir em situações de emergência. Assim, confirma-se mais uma vez que os benefícios decorrentes desses sistemas também apresentam desafios organizacionais, principalmente no âmbito ético.

Por esses motivos, alguns autores desenvolveram uma pesquisa sobre um sistema baseado em auditoria ética (*Ethics-based auditing - EBA*) uma forma de fazer com que os sistemas ADMS estejam de acordo com os princípios éticos¹⁶ e em conformidade com a Iniciativa Global da IEEE — Instituto de Engenheiros Eletricistas e Eletrônicos — sobre ética de sistemas autônomos e inteligentes, uma organização que se dedica a abordar questões éticas relacionadas aos sistemas autônomos e foca em promover discussões e desenvolver diretrizes para garantir que esses sistemas sejam desenvolvidos e utilizados de forma ética e responsável em diversos setores.¹⁷

A “auditoria ética”, refere-se ao processo de avaliar e analisar as práticas, políticas e procedimentos de uma organização ou indivíduo com o objetivo de garantir a conformidade com padrões éticos e morais. Entretanto, antes de descrever um pouco mais sobre esse processo, é importante ressaltar que a busca por formas de garantir que sistemas autônomos operem de maneira ética não é recente e vem de histórias de ficção científica, como o conto “Andando em Círculos” ou “Runaround” de Asimov,¹⁸ em que o autor estabelece as famosas “três leis da robótica”, que ficaram notórias na cultura pop e nos cinemas. Leis que priorizam a segurança humana, proibindo que um robô cause dano a um ser humano ou permita que sofra algum mal, demandam a obediência dos robôs às ordens humanas, exceto quando entram em conflito com a Primeira Lei, e instruem os robôs a protegerem sua própria existência, desde que isso não entre em conflito com as duas primeiras leis. Tais leis em muitos momentos podem ter servido de inspiração para a criação de medidas para regularizar o uso e desenvolvimento de IA.

Além disso, existem grupos de pesquisa de relevância sobre o assunto, como o de especialistas de alto nível em Inteligência Artificial convocado pela Comissão Europeia para fornecer orientações e recomendações sobre o desenvolvimento e a implementação de

¹⁵ Pinheiro (2020).

¹⁶ Mokander et al. (2021).

¹⁷ Tal assunto é discutido por Chatila e Havens (2019).

¹⁸ Asimov (2015).

políticas e estratégias relacionadas à IA na União Europeia.¹⁹ As recomendações do AI HLEG são importantes para orientar a legislação e regulamentação relacionadas à IA na União Europeia. Evoluindo os debates em relação a ética do o que precisa estar no radar quando se trata da IA e para como a ética pode ser aplicada de forma eficaz em situações reais.

A orientação resultou na proposta legislativa conhecida como “EU Artificial Intelligence Act” (Lei de Inteligência Artificial da União Europeia), apresentada pela Comissão Europeia em abril de 2021.²⁰ Essa legislação tem o objetivo de regular o desenvolvimento, a implementação e o uso de sistemas de inteligência artificial dentro dos países membros da União Europeia.

A proposta abrange uma série de disposições, incluindo exigências de transparência, requisitos de documentação, avaliações de risco, e a proibição de determinadas práticas de IA consideradas de alto risco. Também estabelece a criação de uma agência reguladora para supervisionar a conformidade com as regras, entretanto não indica de forma abrangente e clara como os instrumentos podem ser utilizados para salvaguardar a transparência e a responsabilização.

É relevante observar que esses valores não se limitam aos Estados-membros da UE e são partilhados por sociedades além das fronteiras da UE. No Brasil, a regulamentação sobre IA ainda está em fase de desenvolvimento e discussão, não possuindo uma legislação específica até a data de conclusão desta pesquisa. No entanto, já existem conselhos e debates sobre Regulação de IA e projetos de lei que dispõem sobre o seu uso, como o Projeto de Lei nº 2338 de 2023, do [Congresso-Nacional \(2023\)](#).

Os mecanismos de governança previamente expostos para garantir a ética na IA, trazem junto consigo questões práticas que podem possuir diversos tipos de focos iluminados. [Mokander et al. \(2021\)](#) tratam justamente da aplicação desse conhecimento nos processos de desenvolvimento de *softwares*, a promoção da diversidade nas equipes de desenvolvimento de *softwares*, processo de avaliar e analisar os dados de entrada em um sistema de inteligência artificial para identificar possíveis vieses, verificar os mecanismos de decisão nos modelos e códigos, à avaliação ou análise dos possíveis efeitos e consequências de uma determinada ação e a consideração dos contextos de operação de sistemas ADMS.

No contexto apresentado por [Floridi \(2019\)](#), há um caminho mais prático para a integração da ética na IA. Esse enfoque ressalta alguns dos riscos associados:

1. Compras éticas (*ethics shopping*): Se dispõe sobre a prática questionável de selecionar princípios éticos a fim de justificá-los retroativamente, em vez de implementar ou aprimorar novos comportamentos por meio da comparação com padrões éticos de

¹⁹ [HIGH-LEVEL... \(2023\)](#).

²⁰ [Madiega \(2021\)](#).

domínio público [Floridi \(2019\)](#). O que causa confusão e inconsistência entre as partes interessadas,

2. Lavagem ética (*ethics bluewashing*): Refere-se à prática indevida de empresas ou organizações que tentam criar uma imagem de responsabilidade ética e ambiental, fazendo afirmações infundadas ou enganosas sem efetivamente implementar políticas ou práticas.
3. *Lobbying* ético (*ethics lobbying*): São os esforços de grupos de interesse ou empresas para influenciar políticas públicas e decisões governamentais, com a auto-regulação sobre a ética da IA. Na prática de explorar a ética digital para atrasar, reverter, substituir ou evitar legislação necessária sobre a concepção
4. Despejo ético (*ethics dumping*): Expressão cunhada em 2013 pela Comissão Europeia para descrever a prática de transferir atividades, como pesquisa ou produção, para locais onde as regulamentações éticas e padrões de responsabilidade são mais fracos ou inexistentes.
5. Esquiva à ética (*ethics shirking*): Consiste na prática ilícita de aplicar padrões duplos nas avaliações morais, em evitar ou contornar questões éticas deliberadamente, muitas vezes ignorando responsabilidades éticas ou tomando medidas para evitar a prestação de contas por ações que possam ser consideradas antiéticas.

Sendo assim, há a perspectiva de reduzir as práticas descritas e analisadas no artigo de [Floridi \(2019\)](#) na medida em que se basearem na compreensão de que esses “atalhos” não tornam o caminho para soluções éticas genuínas mais fáceis, mas, pelo contrário, complicam ainda mais os problemas existentes em práticas em ética digital.

Para lidar com essas preocupações, o projeto de lei de IA da regulamentação europeia segue a abordagem de análise baseada em risco,²¹ na qual a intervenção legal é adaptada ao nível de risco de opacidade, complexidade, dependência de dados e de como o comportamento autônomo pode afetar adversamente vários direitos fundamentais e a segurança dos usuários. A fim de tornar claro os contextos nos quais diferentes formas de avaliação de riscos são necessárias na aplicação da IA traz consigo impactos éticos diversos em relação ao risco do uso de tipos gerais.

Por isso, a IA tem que ser implementada conforme regulamentações e estratégias para dispositivos planejados de maneiras variadas. Para reduzir esses riscos lidando de forma adequada a cada nível [Madiega \(2021\)](#) distingue sistemas com risco inaceitável, alto risco, risco limitado e baixo ou mínimo risco. Dentre cada uma dessas categorias, são propostas medidas regulatórias específicas como pode-se verificar na Figura 5 a representação, que se baseia na pirâmide de riscos da IA.

²¹ [Madiega \(2021\)](#).

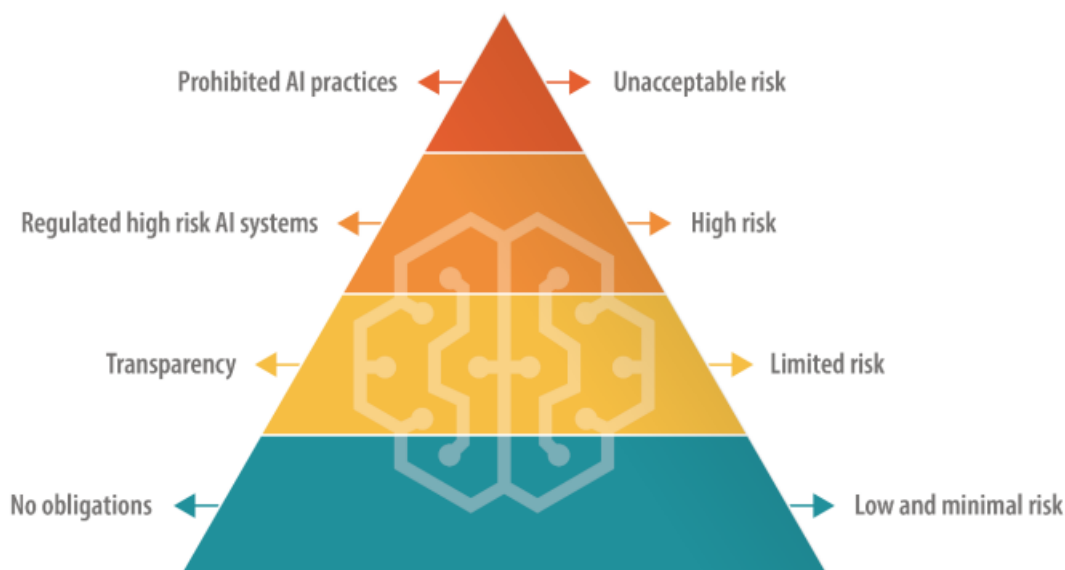


Figura 5 – Pirâmide de riscos. Fonte: [Madiega \(2021\)](#).

Dado que as aplicações de IA seriam regulamentadas apenas na medida do necessário para lidar com níveis de risco, é importante descrever a ideia geral de cada caso de risco, tendo em mente a preservação dos direitos fundamentais [Unicef \(2023\)](#) e o que está definido no [Madiega \(2021\)](#). No entanto, é importante notar que a avaliação de risco pode variar dependendo do contexto e da aplicação específica da IA.

No caso de risco inaceitável, são proibidas práticas de IA prejudiciais que representem uma ameaça clara à segurança, meios de subsistência e direitos das pessoas, como a utilização de técnicas manipuladoras subliminares e a exploração de grupos vulneráveis.

Para os sistemas de alto risco, que podem ter impactos adversos na segurança e nos direitos fundamentais, são estabelecidas regulamentações específicas. Por exemplo, um sistema autônomo de veículo que apresenta falhas frequentes na detecção de obstáculos e tomada de decisões de direção. Esses sistemas podem ser empregados como componentes de segurança de produtos ou como produtos sujeitos à legislação de harmonização de saúde e segurança da UE. Neste caso, se incluem sistemas de alto risco como identificação biométrica e categorização de pessoas naturais, gestão e operação de infraestrutura crítica, entre outras.²²

Para sistemas de risco limitado, refere-se a situações em que o potencial de impactos adversos na segurança e direitos fundamentais das pessoas é considerado relativamente baixo. Esses sistemas têm menor probabilidade de causar danos significativos ou violar direitos fundamentais, como aplicativos de previsão do tempo, tradutores automáticos de idiomas e assistentes virtuais de compras online. Por isso, o nível de regulamentação é ajustado de acordo com o nível de risco apresentado.

²² [Madiega](#).

Por fim, para sistemas de baixo ou mínimo risco, menos propensos a representar ameaças sérias à segurança e aos direitos das pessoas em comparação com sistemas classificados como de “risco alto” ou “risco inaceitável”. Neste caso se enquadram: filtros de spam de E-mail e sistemas de recomendação de conteúdo em plataformas de streaming. Assim, não há obrigações legais adicionais, permitindo o desenvolvimento e uso na UE sem restrições especiais.

Mesmo assim, existe uma incerteza sobre os papéis e responsabilidades dos desenvolvedores, provedores e usuários de sistemas de IA. Ao aplicar o conceito de “ética da informação como ética da computação em rede”, de²³ torna-se evidente que a definição de riscos e responsabilidades na cadeia de valor da IA deve ser fundamentada no impacto real e potencial sobre indivíduos e na sociedade como um todo, ou seja, os de “alto risco”, principalmente, devem se basear no dano mensurável e no impacto potencial.

Uma perspectiva complementar para analisar os requisitos regulatórios relacionados a diferentes aplicações e riscos de Sistemas de Tomada de Decisão Automatizada é proposta por Krafft, Zweig e König (2022). Os autores propõem uma análise dos requisitos regulatórios, considerando as diversas aplicações e os riscos associados a esses sistemas que se difere da descrição de riscos da UE.²⁴ Essa perspectiva complementar oferece uma visão abrangente das medidas necessárias para garantir a responsabilidade e a segurança na implementação dos ADMs em uma variedade de contextos como pode ser distinguido na matriz de risco da figura 6.

Conforme representado na figura, as cores mais próximas do vermelho representam o maior risco, enquanto as de verde representam os cenários mais seguros. Ele tem o impacto analisado de “*scope of agency loss*”, que se refere à extensão das perdas potenciais que podem ocorrer devido aos conflitos de interesses entre o principal e o agente em relação à intensidade potencial de perigo.

Seguindo esta ordem, tem-se: (4) Sistemas sem administração; (3) responsabilidade no processo, que se refere à responsabilidade pelo processo em um contexto organizacional ou operacional, onde os indivíduos ou entidades envolvidos são considerados responsáveis por garantir que ele seja conduzido de maneira eficaz, eficiente e de acordo com os padrões e regulamentos estabelecidos; (2) responsabilidade baseada em objetivos, que se refere à avaliação da IA com base em metas e objetivos específicos, em vez de apenas em métricas tradicionais de desempenho; (1) responsabilidade pelos resultados, onde a IA é julgada com base em sua capacidade de alcançar objetivos definidos, em vez de apenas medir métricas de precisão ou acurácia (especialmente necessárias em sistemas de alto risco da UE); (0) “responsabilidade de marketing”, que implica o uso de algoritmos e técnicas analíticas para quantificar e avaliar o impacto das estratégias de marketing em termos de

²³ Floridi (2019).

²⁴ Madiaga (2021).

resultados financeiros.²⁵

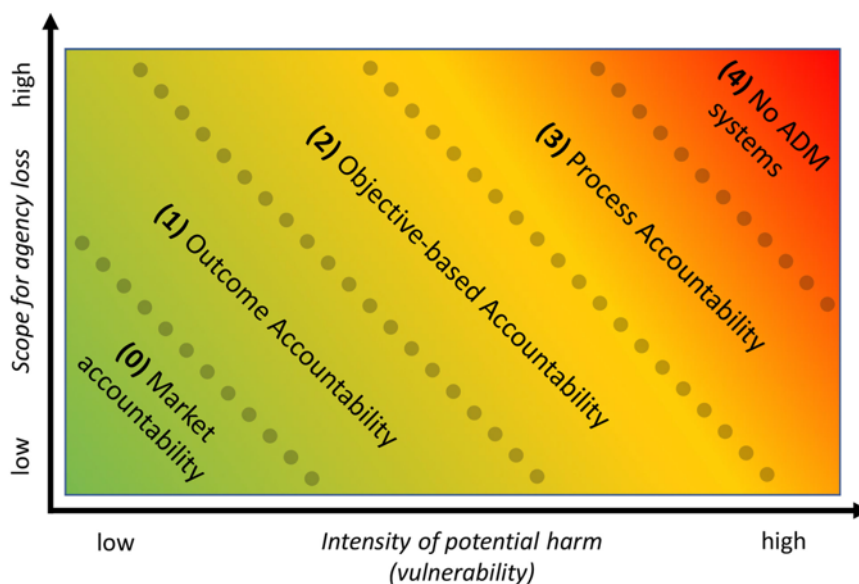


Figura 6 – Matriz de risco. Fonte: Krafft, Zweig e König (2022).

O terceiro aspecto destacado no “manifesto” tem impacto na vida humana causado pela transição digital, “... Mudando de escassez para abundância, no que diz respeito à informação”, relaciona-se de forma direta com as definições de risco da IA da UE,²⁶ uma vez que um grande volume de dados disponíveis pode tornar os algoritmos mais complexos e, conseqüentemente, mais difíceis de prever para os desenvolvedores compreenderem completamente como o sistema toma decisões específicas em situações particulares. Isso pode gerar dificuldades na interpretação do processo de tomada de decisão, o que é crucial para entender o comportamento e avaliar os riscos associados.

Além disso, a opacidade também pode ser agravada pela presença de dados irrelevantes ou contraditórios no conjunto de dados. Isso pode confundir o sistema e levar a decisões imprevisíveis ou inadequadas em certas circunstâncias.

Por fim, a falta de transparência nas técnicas de aprendizado de máquina utilizadas nos sistemas de IA pode contribuir para a opacidade. Algoritmos de aprendizado profundo, por exemplo, são notoriamente difíceis de interpretar, o que torna desafiador compreender o raciocínio por trás das decisões tomadas pelo sistema.

Essa complexidade pode influenciar diretamente a avaliação de risco dos sistemas de IA. Esses fatores combinados podem resultar em um sistema de IA que opera de maneira opaca, ou seja, cujo funcionamento interno e processo de tomada de decisão são

²⁵ Observa-se dessa forma que as duas descrições de risco, da UE citada em [Madiega \(2021\)](#) e do artigo [Krafft, Zweig e König \(2022\)](#), podem ser analisadas complementarmente, e que refletem alguns aspectos do contexto apresentado por [Floridi \(2019\)](#) sobre os riscos associados à falta de ética na prática da Ética Digital.

²⁶ [Madiega \(2021\)](#).

difíceis de compreender ou explicar completamente, como uma caixa-preta, o que aumenta a necessidade de medidas que garantam a responsabilidade e a transparência.

De outra forma, a interconexão extensiva nas relações sociais também está em jogo. Com a proliferação de informações e a crescente interconexão entre diferentes sistemas e atores sociais, a influência da IA nas relações sociais é cada vez mais significativa. Isso significa que as decisões tomadas por sistemas de IA têm o potencial de afetar uma ampla gama de interações sociais, desde recomendações de conteúdo até decisões de contratação em ambientes profissionais, como explicitado anteriormente.

Portanto, mais uma vez se mostra crucial estudar formas de salvaguardar a transparência e a responsabilização na implementação e regulamentação dos sistemas de IA, para que mesmo em um cenário de abundância de informações e complexidade algorítmica, as implicações sociais e os riscos associados sejam devidamente considerados. [The ethics of artificial intelligence \(2014\)](#) destaca a importância dos algoritmos serem robustos contra manipulação e que mesmo quando sistemas de IA são projetados por organizações, seja possível encontrar a pessoa, ou o grupo, responsável caso algo dê errado.

A transparência possibilita que os usuários compreendam o processo decisório, enquanto a responsabilidade implica que os desenvolvedores e operadores de sistemas de IA se responsabilizem pelos resultados de suas criações. Conforme [Alves e Andrade \(2021\)](#) sugerem, uma boa prática para o desenvolvimento de sistemas transparentes é a elaboração de IA explicável (XAI), que fornece informações para que ocorra uma compreensão qualitativa sobre o seu processo de predição. Desta maneira, busca-se transformar a “caixa-preta” algorítmica em uma “caixa de vidro”, permitindo verificabilidade por meio de auditoria ou técnicas de avaliação.

Tais práticas, como a técnica EBA discutida por [Mokander et al. \(2021\)](#), visam identificar correlações e vieses, promover rastreabilidade e, conseqüentemente, aumentar a confiança dos usuários e da sociedade na IA. Ao explorar essas contribuições técnicas, torna-se evidente que a busca pela transparência nos ADM é um campo em constante evolução, com pesquisadores e profissionais engajados em encontrar soluções práticas para um uso mais responsável e confiável desses sistemas.

Refirmando a argumentação, uma abordagem adicional para lidar com a complexidade dos modelos de aprendizado de máquina utilizados no desenvolvimento de IA, especialmente em sistemas de tomada de decisão de alto risco em áreas como saúde e justiça, foi discutida no artigo de [Rudin \(2019\)](#). Essa abordagem propõe a redução da opacidade algorítmica, sugerindo que uma solução viável seria adotar modelos interpretáveis. Isso ressalta a percepção de que existem várias tentativas de explicar esses modelos opacos, mas o enfoque deveria ser direcionado para a criação de modelos interpretáveis, a fim de garantir que o progresso contínuo dessas tecnologias não resulte em danos catastróficos a longo prazo.

O autor aponta ainda que os métodos de explicabilidade muitas vezes não fornecem explicações fiéis ao que o modelo original calcula. Mesmo quando corretos, os métodos de explicabilidade podem deixar de fora informações essenciais. Para reforçar a ideia, o autor expõe o mito do trade-off, que em aprendizado de máquina refere-se à crença comum de que há um compromisso inevitável entre a precisão (quão bem o modelo faz previsões) e a interpretabilidade (quão compreensível ou explicável é o modelo). No entanto, o autor destaca que esse dilema não é absoluto. Em alguns casos, especialmente quando lida-se com dados estruturados e características significativas, modelos mais simples e interpretáveis podem alcançar um desempenho preditivo tão bom quanto (ou até melhor do que) os modelos mais complexos. Portanto, o *trade-off* entre precisão e interpretabilidade não é uma regra universal, e em muitos cenários, é possível ter ambos.

Essa perspectiva desafia a ideia preconcebida de que precisão e interpretabilidade são mutuamente exclusivas, sugerindo que é possível buscar modelos que sejam precisos e compreensíveis ao mesmo tempo. Também destacam-se as limitações dos métodos de explicabilidade utilizados para tornar compreensíveis os modelos de “caixa preta”.²⁷ Esses métodos podem falhar em fornecer explicações abrangentes e precisas, mesmo quando corretos, pois podem omitir informações essenciais sobre o funcionamento do modelo original. O autor ilustra esse ponto com o exemplo de mapas de saliência em processamento de imagens.

Os mapas de saliência são uma técnica de interpretabilidade frequentemente usada em visão computacional, especialmente em modelos de aprendizado profundo. Esses mapas ajudam a entender quais partes da entrada (como uma imagem) são mais influentes para a decisão do modelo. Eles funcionam destacando visualmente as áreas da entrada que mais contribuem para a predição feita pelo modelo. Isso é feito atribuindo valores a cada pixel na imagem, indicando a importância relativa de cada pixel para a decisão final.

Além disso, os mapas de saliência têm suas limitações.²⁸ Embora esses mapas indiquem onde o modelo está convergindo, eles não necessariamente explicam como o modelo está utilizando essas informações para fazer a predição. Além disso, em alguns casos, eles podem não fornecer uma compreensão completa ou precisa do processo de decisão do modelo. Portanto, enquanto os mapas de saliência são uma ferramenta valiosa para interpretabilidade, é importante reconhecer que eles oferecem uma visão limitada e podem não revelar todos os detalhes sobre como o modelo opera internamente.

Uma das alternativas para esses casos seria a implementação do sistema de Inteligência Artificial Explicável (XAI) desenvolvido pela Agência de Projetos de Pesquisa Avançada de Defesa (DARPA) dos Estados Unidos.²⁹ A DARPA está interessada em criar

²⁷ Rudin.

²⁸ Rudin.

²⁹ Gunning (2016)

tecnologia para tornar essa nova geração de sistemas de IA explicável, especialmente para o Departamento de Defesa, que precisa de um sistema cuja eficácia não seja limitada pela incapacidade das máquinas de explicar suas decisões e ações aos usuários humanos.

De maneira simplificada o sistema XAI tem como alvo um usuário final que depende de decisões,³⁰ recomendações ou ações produzidas por um sistema de IA e, portanto, precisa entender a justificativa para as decisões do sistema, e isso se dá focando em 3 ênfases principais, que se dá conforme a Figura 7.



Figura 7 – Conceito XAI adaptado de: [Gunning \(2016\)](#).

Os desenvolvedores da metodologia XAI se propõem a produzir modelos mais explicáveis, projetar a interface de explicação e entender os requisitos psicológicos para explicações eficazes. Com uma explicação mais visual como a Figura 8, [Gunning \(2016\)](#) apresentam uma comparação entre os mecanismos de ML mais comuns hoje e como ocorre na XAI, ressalta principalmente o fato de que a XAI busca fornecer explicações para o usuário, para que ele entenda como o sistema chegou a indicação ou solução apresentada.

[Rudin \(2019\)](#) também discute os principais desafios associados à interpretabilidade em modelos de aprendizado de máquina. Destaca que, em muitos casos, mesmo havendo preferência por modelos interpretáveis, são utilizadas caixas pretas com explicações, principalmente devido a questões relacionadas a lucro e propriedade intelectual. Além disso, ressalta que a construção de modelos interpretáveis demanda esforço computacional e conhecimento no domínio específico da aplicação. Também menciona que modelos de caixa preta são frequentemente vistos como capazes de descobrir padrões ocultos nos dados, embora modelos interpretáveis também possam realizar essa tarefa se forem adequadamente projetados.

³⁰ [Gunning](#).

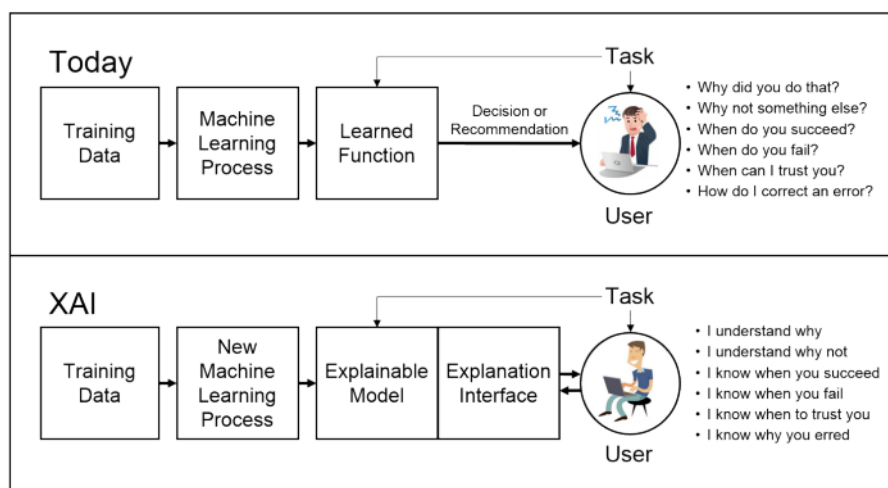


Figura 8 – Comparação XAI. Fonte: [Gunning \(2016\)](#).

O autor ainda destaca a existência de uma grande quantidade de pesquisas sobre interpretabilidade que muitas vezes não são explicitamente rotuladas como tal, o que pode dificultar a revisão abrangente desses estudos. Além disso, questiona a necessidade de revisões sobre interpretabilidade, argumentando que ela é tão específica quanto as medidas de desempenho, e que interpretabilidade e explicabilidade são conceitos distintos.

Ao argumentar sobre a importância de existirem modelos interpretáveis em vez de explicáveis, demonstra-se a necessidade de incentivar a formulação de políticas que busquem soluções nesse sentido, como esforços na garantia de segurança e confiança em modelos de aprendizado de máquina para decisões de alto risco.

Da mesma forma, o [Krafft, Zweig e König \(2022\)](#) sublinham a importância crítica de adotar uma abordagem regulatória diferenciada para ADM. Ao considerar a diversidade de aplicações e os variados impactos associados a esses sistemas, torna-se evidente que uma única abordagem regulatória não seria suficiente para abranger todas as situações. Portanto, é essencial que a regulamentação precisa seja adaptável, de forma a evitar restrições excessivas em algumas áreas e garantir a segurança em outras.

3.3 Estratégias de Governança Ética na Era da IA

Tendo isso em mente, mostra-se necessário a existência de mecanismos que tenham o propósito de detectar, avaliar, reconhecer e mitigar esses riscos em aplicações reais de sistemas de ADM, ou seja, mecanismos reguladores que garantam a responsabilidade ética na IA. Como a IA pode ser aplicada de maneira variada, não existem soluções fixas que resolvam todos os problemas, por isso, há uma enorme quantidade de pesquisa em desenvolvimento para produzir informações relevantes que facilitem a aplicação ética dos sistemas desenvolvidos, não para aprender a como codificar a ética.

Dentre esses estudos, destacam-se aqueles que propõem desenvolver um mecanismo de governança inspirado nas auditorias que já ocorrem em companhias,³¹ organizações e empresas, mas desta vez focado em atender as demandas éticas ocasionadas pelas formas em que sistemas de IA são desenvolvidos e implantados na sociedade na EBA.

Isso ressalta que a solução para produção de sistemas eticamente engajados não se dá apenas pela estruturação e publicação de um código de conduta, como o desenvolvido pela UE,³² ou apenas uma certificação da própria organização desenvolvedora de que a IA está de acordo com os padrões. A EBA se trata de um processo estruturado para verificar as implicações atuais e futuras, além das de construção de um sistema de ADMS, com o propósito de avaliar de maneira distinta para cada objetivo ou tipo de sistema auditado se eles atuam de acordo com as normas. Dessa forma, esse sistema de auditoria também deve ocorrer com auditores independentes da organização auditada, e pode avaliar a estrutura de diversas formas, dentre elas: a pessoa desenvolvedora, a organização ou o sistema técnico.

Na Figura 9 pode-se observar um mapa de como se dão as relações entre as organizações que desenvolvem os sistemas ADMS, reguladores, e os auditores independentes.

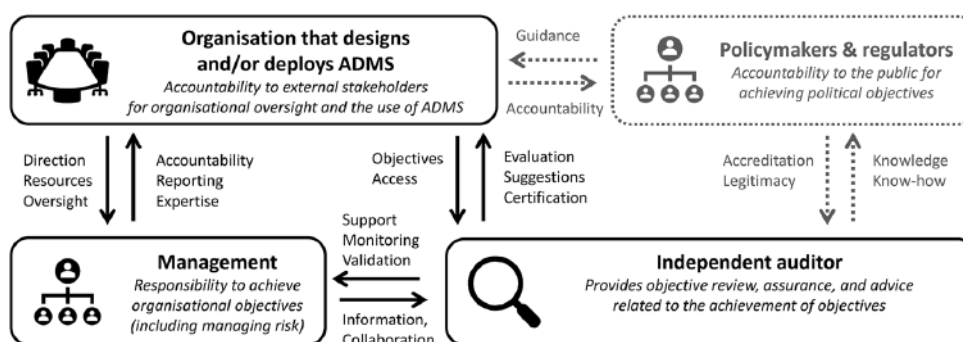


Figura 9 – Funções e responsabilidades durante auditorias independentes. Fonte: Mokander et al. (2021)

A auditoria é sempre direcionada a um propósito específico, no caso da EBA, garantir que os ADMS operem de acordo com diretrizes éticas. A documentação gerada no processo permite identificar por que um ADMS estava errado, ajudando a prever consequências indesejadas e evitar erros futuros. A auditoria pressupõe independência operacional entre o auditor e o auditado para evitar conflitos de interesse e esclarecer responsabilidades por diferentes tipos de danos ou falhas no sistema. Para ser eficaz, os auditores devem testar os ADMS em uma ampla variedade de cenários. Reguladores podem apoiar a implementação de procedimentos de EBA fornecendo a infraestrutura para compartilhamento de informações e criação de relatórios padronizados.

³¹ Mokander et al. (2021).

³² Madiega (2021).

Entretanto, é importante que exista uma avaliação holística desses “lugares”, o que proporciona maior confiabilidade em relação à funcionalidade do sistema, aos impactos, à segurança e à transparência do uso de dados. Com um processo claro, as iniciativas EBA de *frameworks* e de ferramentas que existem atualmente podem ter duas origens distintas:³³:

1. A origem de estratégias nacional ou regional, como *SMART Dubai*,³⁴ que se ancoram nos aspectos legais ou guias de normas,³⁵ o que proporciona ferramentas para ajudar os designers a entender as falhas do sistema.
2. Uma estratégia de “abotoaR”, que significa que se transferem os princípios éticos para mecanismos de governança, com uma contextualização, gerenciamento de dados e uso de ADMS. Uma desvantagem deste caso é que geralmente não se faz uma avaliação holística, e acaba focando em alguma das preocupações principais como privacidade e isso pode levar ao risco de se ter uma boa expectativa em relação ao seu funcionamento com base neste setor, mas sem saber ao certo em que grau está em relação ao todo. Destaca-se aqui a possibilidade de existir um sistema interno de auditoria ética dentro das organizações que possibilitem um acompanhamento independente das normas dentro da companhia.

A existência desses *frameworks* se dão a partir de diferentes ferramentas de medidas, avaliação e visualizações das propriedades dos sistemas de IA, que também já vêm sendo desenvolvidas por companhias privadas e pesquisas acadêmicas. Um dos exemplos é a ferramenta FAIRVIS, que permite visualizar as saídas do sistema e a análise para descobrir preconceitos interseccionais em aprendizado de máquina.³⁶ As ferramentas foram desenvolvidas para ajudar a tornar a EBA mais democrática e podem se ancorar em explicações baseadas em dados, em modelos da base algorítmica do próprio sistema, ou em explicações baseadas na comparação do propósito do sistema com seus resultados.

De maneira geral, a visão para Auditoria Ética de Sistemas de Tomada de Decisão Automatizada envolve a integração de estruturas e ferramentas existentes de Auditoria Ética em um processo estruturado que monitora cada etapa do ciclo de vida do desenvolvimento de *softwares*. Essa abordagem combinada inclui auditoria de funcionalidade, que avalia a lógica por trás das decisões, auditoria de código, que revisa o código-fonte do algoritmo, e auditoria de impacto, que investiga os efeitos das saídas do algoritmo. O estudo de Floridi também destaca vantagens fundamentais para o efetivo gerenciamento dos riscos éticos associados aos ADMS: EBA pode fornecer suporte à tomada de decisão, aumentar

³³ Mokander et al.

³⁴ Salem (2016).

³⁵ Madiaga (2021).

³⁶ Cabrera et al. (2019).

a confiança pública na tecnologia, alinhar ética e legislação de forma contextualizada, antecipar e mitigar potenciais consequências adversas, equilibrar conflitos de interesse e facilitar a alocação de responsabilidades. Se os *frameworks* estiverem de acordo com os critérios descritos pelos autores que envolvem.

A EBA emerge como uma ferramenta crucial no gerenciamento dos riscos éticos associados aos ADMS. Suas vantagens metodológicas, quando devidamente implementadas, fornecem suporte à tomada de decisão, promovem transparência e confiança pública, permitem a adaptação contextualizada de ética e legislação, antecipam e mitigam consequências adversas, equilibram interesses diversos e facilitam a alocação de responsabilidades. Contudo, a plena eficácia desses benefícios depende da concepção e implementação adequada dos *frameworks* de EBA, da perspectiva dos *stakeholders* sobre a EBA de ADMS, assim como o design das práticas de EBA. Embora esses critérios sejam aspiracionais e possam ser desafiadores de serem plenamente atendidos, representam um guia valioso para tomadores de decisão e organizações na concepção e implementação de procedimentos EBA.

Em resumo, as práticas de EBA, mesmo que implementadas de forma imperfeita, têm o potencial de causar impacto significativo na maneira como os ADMS são desenvolvidos e implantados. Embora os critérios propostos sejam ambiciosos e possam ser desafiadores de serem totalmente cumpridos, é essencial não deixar a busca pela perfeição ser um obstáculo para o progresso em relação a implementar iniciativas para reduzir o impacto ético das aplicações de IA. Assim, políticos e organizações envolvidas no design e implementação de ADMS são encorajados a considerar esses sete critérios ao desenvolver e implementar procedimentos de EBA.

Entretanto, as aplicações de sistemas EBA trazem consigo desafios financeiros e sociais, pois impõem custos financeiros que podem ser desproporcionais em empresas menores. Dessa forma, a sociedade pode enfrentar questões sobre quais *stakeholders* colheriam quais benefícios e arcariam com quais custos. Assim, o ponto de [Mokander et al. \(2021\)](#) é que tanto os custos quanto os benefícios associados à EBA devem ser distribuídos de forma a não sobrecarregar ou beneficiar indevidamente grupos específicos ou setores da sociedade. Alinhando demandas éticas equilibradas com incentivos para inovação e adoção para garantir que os desenvolvimentos sejam robustos e éticos no progresso tecnológico. Em relação a esses sistemas de auditoria, outro risco é que as empresas burlarem os sistemas para entregar o resultado esperado, enquanto na verdade não estão operando da forma correta, um risco também destacado quando falamos sobre a necessidade de transparência.³⁷

Também existe o risco de ocorrer o “*nudging*”, o processo de influenciar preferências pessoais por meio de reforço positivo ou sugestão indireta, o que pode deslocar a linha de base normativa contra a qual o alinhamento ético é avaliado. Esse risco é agravado pelo

³⁷ [Alves e Andrade \(2021\)](#).

'viés de automação', ou seja, a tendência das pessoas a confiar em informações provenientes de máquinas mais do que em seu próprio julgamento. Neste ponto, encontra-se o quarto aspecto profundamente afetado na vida humana destacado [The Onlife Manifesto \(2015\)](#), no seu quarto item, ao passar da primazia das entidades sobre as interações para a primazia das interações sobre as entidades. Essa faceta afetada também se dá em outros processos mais comuns no dia a dia do ser humano, por meio de notícias ou notícias falsas (*fake news*), não notadas por sistemas de regulação ou ferramentas que corroboram em manter o indivíduo conectado com as redes sociais, como a utilizada pelo *Instagram* e *TikTok* ao recomendar ainda mais assuntos correlacionados ao interesse do cidadão, independente da veracidade das informações ou do quanto elas são enviesadas em manter comportamentos antiéticos se proliferando na sociedade.

Sobre esse assunto, [Santaella e Kaufman \(2021\)](#) fazem a distinção entre os neologismos “datificação” e “dataísmo”. Datificação é a tendência contemporânea de transformar praticamente todos os aspectos da vida em dados. Isso implica uma visão positiva da coleta de dados como algo valioso para a compreensão e melhoria do mundo. Por outro lado, o *dataísmo* é uma postura filosófica que sugere que os dados e os algoritmos que os manipulam têm autoridade suprema na tomada de decisões, até mesmo sobre questões éticas e morais, muitas vezes relegando a intuição humana a um segundo plano.

Tendo em vista o dataísmo, [Santaella e Kaufman \(2021\)](#) refletem que um seguidor dessa filosofia acredita na liberdade de informação e confia mais no *big data* e nos algoritmos de computador do que no conhecimento e sabedoria humanos. O que, de forma psicológica, pode trazer prejuízos à capacidade mental humana sobre a interação com o mundo exterior,³⁸ conforme expresso no livro “Re-engineering humanity”, que explica sobre o risco de deixar o comportamento dos seres humanos programáveis e previsíveis, transformando substancialmente a capacidade do ser em produzir a si mesmo e suas escolhas. Por outro lado, nas redes sociais, como destacam [Santaella e Kaufman \(2021\)](#), a produção de “gêmeos digitais” do ser humano. No meio corporativo, a expressão “*digital twins*” se define como modelos de simulação de produtos, mas que agora podem ser vistos na *Alexa*, assistente da *Amazon*.

No contexto atual, em que a inteligência artificial desempenha um papel cada vez mais relevante na sociedade, a EBA emerge como uma ferramenta fundamental para garantir que os sistemas de tomada de decisão automatizada sejam desenvolvidos e implantados de maneira ética e responsável. A abordagem abrangente da EBA, que inclui auditorias de funcionalidade, de código e de impacto, oferece uma estrutura robusta para avaliar e mitigar os riscos éticos associados aos ADMS

No entanto, para que a EBA atinja todo o seu potencial, é essencial que os custos e Benefícios sejam distribuídos de forma justa na sociedade, evitando sobrecarregar grupos

³⁸ [Frischmann e Selinger \(2018\)](#).

específicos. Além disso, a crescente influência dos dados e algoritmos na tomada de decisões, conforme destacado pelo conceito de dataísmo, traz à tona questões fundamentais sobre a autonomia humana e a necessidade de equilibrar a tecnologia com a sabedoria e o conhecimento humanos.

Outra forma necessária de se analisar a Ética na IA é por meio da ética profissional, como [Glock e Goldim \(2003\)](#) destacam como mais que uma simples reflexão, mas um compromisso social de todos os profissionais. Neste estudo, conforme proposto nos objetivos do artigo aprofundaremos essa reflexão no contexto de profissões que atuam diretamente na produção de IA, como engenharia da computação, controle e automação e áreas correlatas.

Conforme destacado por [Glock e Goldim \(2003\)](#), na formação profissional que engloba a aquisição de competências e habilidades específicas da área de atuação, é crucial integrar a reflexão desde o início dos estágios práticos. Ao concluir o ensino superior, o profissional faz um juramento, que implica voluntariamente seguir um conjunto de normas consideradas ideais para a prática da profissão. Isso representa a base moral da Ética Profissional, que envolve aderir a um conjunto de normas consideradas ideais para o exercício da profissão.

A consideração ética contribui para o desenvolvimento sustentável e equitativo da sociedade e previne práticas prejudiciais ou injustas, o desconhecimento pode levar a processos ético-disciplinares nos conselhos profissionais. Portanto, promover e cultivar uma consciência ética deve ser uma prioridade para todos os profissionais, especialmente aqueles que têm um impacto tão significativo nas vidas das pessoas como os que atuam com IA.

3.4 Trilhando o Caminho Ético

Dentro do mesmo tema, torna-se preciso compartilhar a percepção dos desafios de [Piteira, Aparicio e Costa \(2019\)](#), que abordam o código de ética e conduta profissional da ACM,³⁹ e o *IEEE Ethical Design*,⁴⁰ fornecendo sugestões de orientações para uma conduta profissional ética na área da computação. Entretanto, sabe-se que as práticas mudam, que outros pontos de vista emergem e o conhecimento cresce.

O Código de Ética e Conduta Profissional da ACM (Association for Computing Machinery),⁴¹ e IEEE são duas organizações proeminentes no campo da tecnologia e da computação, ambas desempenhando papéis cruciais no estabelecimento de padrões éticos e profissionais para os profissionais dessas áreas. Ambas têm códigos de ética que enfatizam a integridade, responsabilidade e respeito, apesar de existirem diferenças em seus focos. O

³⁹ [ACM.org \(2018\)](#).

⁴⁰ [IEEE... \(2023\)](#).

⁴¹ [ACM.org](#).

IEEE tem um escopo mais amplo, abrangendo não apenas a computação, mas também a engenharia elétrica, eletrônica e afins. Por outro lado, a ACM está mais especificamente centrada na ciência da computação.

Embora tenham essas distinções, ambas compartilham a missão geral de promover práticas éticas e de alta qualidade no campo da tecnologia e da computação, contribuindo assim para o avanço responsável e sustentável dessas disciplinas. O que é um dos motivos da análise dos códigos realizada por [Piteira, Aparicio e Costa \(2019\)](#) ser relevante neste estudo. Em um primeiro momento, ele fornece uma compreensão sólida dos princípios éticos e de conduta profissional delineados pela ACM no contexto da Tecnologia da Informação e Sistemas de Informação.

Na Figura 10 constata-se um resumo do código de ética da ACM, onde se destacam os quatro pilares fundamentais: integridade, compromisso com o bem-estar e segurança públicos, competência técnica e profissional e responsabilidade para com as partes interessadas. Posteriormente, esses pilares se subdividem em princípios gerais de ética, fornecendo princípios orientadores para profissionais no desenvolvimento dos sistemas de informação que podem ser reutilizados pelos desenvolvedores de IA. Essa tabela também demonstra uma compreensão aprofundada das implicações éticas no campo da Tecnologia da Informação, servindo como precedente para a IA. A análise reconhece a ênfase dada à responsabilidade dos profissionais em contribuir para o bem-estar da sociedade e evitar danos, aspectos que podem ser levados em consideração por profissionais autônomos ou empresas ao contratar serviços desses profissionais. Da mesma forma, essa abordagem se alinha com a visão de que a tecnologia deve ser desenvolvida e aplicada considerando os impactos sociais e humanos.

Apesar de fornecer nortes para os para os profissionais do campo, ainda não existe um guia passo a passo sobre como aplicar essa metodologia. Isso ocorre em grande parte devido à natureza variável do desenvolvimento de sistemas, que pode ocorrer de maneiras distintas, mesmo com estruturas de conhecimento como a [Russel \(2013\)](#), que oferecem suporte técnico aos programadores e engenheiros. No entanto, em relação aos objetivos, à estrutura empresarial, ao estilo de desenvolvimento, ao tempo e ao contexto em que o desenvolvimento ocorre, ainda no artigo [Piteira, Aparicio e Costa \(2019\)](#) menciona o processo CARE como uma alternativa para aplicar os princípios éticos da ACM. O processo CARE é uma sigla que representa “considerar, analisar, revisar e avalia”, e sua abordagem iterativa facilita o controle e a correção das partes interessadas em relação às responsabilidades éticas e legais envolvidas.

Conforme exposto anteriormente, o artigo também menciona que o IEEE propõe diretrizes para o alinhamento de práticas de design ético e sustentabilidade em sistemas inteligentes. Suas estipulações destacam a importância da segurança, saúde e bem-estar, evitam conflitos de interesse, promovem a honestidade, rejeitam suborno e buscam o

Dimensões de Atuação		Princípios de ética	Ideias Chave		
Princípios Gerais de Ética	Um profissional dos sistemas e tecnologias da informação deve:	Contribuir para a sociedade e para o seu bem-estar, tendo presente que todas as pessoas são partes interessadas nos sistemas de informação (SI /TI)	Deve ser tido em consideração que todas as pessoas são partes interessadas em SI/TI e que devemos contribuir para a sociedade como um todo A honestidade é uma componente essencial da confiança entre as partes		
		Evitar danos.			
		Ser honesto e confiável			
		Ser justo e não tomar ações de discriminação			
		Respeitar o trabalho que requer a produção de ideias, invenções, trabalho criativo e artefactos			
		Respeitar a privacidade			
		Honrar a Confidencialidade			
		Buscar elevados padrões de qualidade: produto e no trabalho profissional			
		Manter altos os standards de competência profissional, competência e praticar a ética			
		Conhecer e respeitar as regras existentes pertinentes para o trabalho profissional			
Responsabilidade Profissional	Um profissional dos sistemas e tecnologias da informação deve:	Acceptar e providenciar uma revisão profissional do trabalho a realizar ou realizado	Um profissional deve orientar a sua atuação para que crie um impacto positivo na sociedade A computação é um serviço para a sociedade Apoiar de forma consistente o bem comum		
		fornecer avaliações abrangentes e completas dos sistemas de computação e seus impactos, incluindo análises de possíveis riscos.			
		Realizar tarefas apenas em áreas da sua competência			
		Promover a conscientização pública e compreensão da computação, relativas à tecnologia e às suas consequências			
		Aceder a recursos de computação e comunicação somente quando autorizado ou quando for obrigado para o bem público			
		Desenhar e implementar sistemas que sejam robustos e usáveis em segurança			
		Defender, promover e respeitar os princípios do The Code.			
		Trate as violações do código como inconsistentes com a participação da ACM.			
		Conformidade com o <i>The Code</i>			Um profissional deverá conhecer, defender, divulgar, promover os princípios de ética do The Code. Deve identificar violações e tratar da sua resolução junto das entidades respetivas.
		Liderança Profissional		Um profissional de computação, quando actua como líder, deve:	Assegurar que o bem público é a principal preocupação.
Articular, incentivar a aceitação e avaliar o cumprimento das responsabilidades sociais pelos membros da organização ou grupo					
Gerir as pessoas e recursos para promover / melhorar a qualidade de vida no trabalho					
Articular, aplicar e apoiar políticas e processos que reflitam os princípios do The Code					
Criar oportunidades para os membros da organização crescerem como profissionais.					
O profissional deve estar ciente das alterações que forem realizadas nos sistemas ou mesmo a sua descontinuidade.					
Reconhecer e cuidar de sistemas que se integrem na infraestrutura da sociedade					

Figura 10 – Tabela resumo do Código de Ética e Conduta Profissional. Fonte: Adaptado da ACM Piteira, Aparicio e Costa (2019).

aprimoramento contínuo de conhecimentos e habilidades tecnológicas.

Por isso, se destaca a necessidade de aumentar a literacia digital e tecnológica na sociedade e enfatiza o tratamento justo e não discriminatório, que refere-se à capacidade de usar, compreender e criar informações de forma eficaz usando tecnologias digitais. Ao mesmo tempo em que se enfatiza a tecnológica na sociedade e enfatiza o tratamento justo e não discriminatório.

Dessa forma, ambas orientações de certo modo se alinham com a filosofia de Floridi et al. (2022) e propõem orientações que favorecem a empregabilidade no campo da IA e desenvolvimento de *softwares*, sugerindo alternativas para que que os profissionais estejam preparados para lidar com os desafios éticos e técnicos associados. Tanto a ACM quanto a IEEE fornecem orientações éticas que promovem um comportamento responsável, especialmente quando se trata de tecnologias de IA. Isso inclui princípios como transparência, responsabilidade e mitigação de riscos.⁴²

A transparência é uma consideração central em ambas as orientações. Isso implica que os engenheiros e desenvolvedores devem ser claros sobre como os sistemas de IA

⁴² Madiega (2021).

funcionam, quais dados são utilizados e como as decisões são tomadas. Essa transparência é essencial para garantir a prestação de contas e para que os sistemas possam ser compreendidos e auditados, reduzindo assim o risco de viés ou comportamento indesejado.

Essas iniciativas e regulamentações fornecem um conjunto de diretrizes claro para os profissionais de IA e desenvolvedores de *softwares*. Eles devem não apenas buscar a inovação e eficácia técnica, mas também garantir que seus sistemas sejam éticos, transparentes e responsáveis. De maneira geral, é um guia fundamental para garantir que a IA seja desenvolvida e utilizada de maneira responsável, abordando preocupações sobre o desemprego e o fetichismo da tecnologia, enquanto se concentra em promover o bem-estar da sociedade. Isso não apenas beneficia a sociedade como um todo, mas também fortalece a confiança do público nas tecnologias de IA e promove a inovação de maneira ética e sustentável.

Dados alguns princípios de auditoria tratados acima, se torna inevitável a relação que a ética em que ter com as logicas corporativas existentes, para que sejam realmente aplicáveis nos modelos de negócios alinhado com atitudes governamentais que impliquem penalidades a falta de utilização de princípios éticos nos interesses corporativos. Em outras palavras, é necessário evitar que as empresas desfrutem da aparência de ética sem a substância conforme destacados nos riscos de lavagem ética.⁴³

Ao considerar o cenário da ética em IA, é crucial também explorar alternativas aos princípios éticos que têm dominado o campo.⁴⁴ Essa perspectiva aponta para a necessidade de ampliar o escopo da discussão, indo além de questões procedimentais, como transparência, e abordando temas mais fundamentais relacionados ao poder e às políticas das gigantes tecnológicas. Isso implica em repensar o papel da IA na sociedade, levando em conta as implicações para comunidades vulneráveis e ecossistemas.

Ambas as abordagens, tanto a busca por justiça em IA quanto a atenção aos aspectos práticos e mensuráveis, desempenham papéis complementares na evolução da ética em IA. Essas perspectivas oferecem um caminho para ir além dos princípios éticos, envolvendo diversos *stakeholders*, como designers, desenvolvedores, gestores e líderes comunitários, na busca por soluções mais concretas e aplicáveis para os desafios éticos na era da IA.

⁴³ Floridi (2019).

⁴⁴ Munn (2023).

4 Considerações Finais

Ao longo deste trabalho, foi explorado a interseção crucial entre ética, Inteligência Artificial e engenharia, destacando a relevância do pensamento de Luciano Floridi nessa interseção. A análise ressalta que a tecnologia é moldada por relações sociais e, no primeiro momento, podem refletir os interesses e valores das elites que a influenciam, sendo, portanto, uma extensão da racionalidade política.

Ao reconhecer a importância desses princípios na governança da IA e na engenharia, compreende-se que, apesar do avanço tecnológico e das IA representarem um progresso para a sociedade, é fundamental compreender melhor a necessidade de abordagens que fortaleçam o comportamento ético de todo o ecossistema tecnológico. Isso abrange instituições governamentais, empresas, desenvolvedores e engenheiros envolvidos na criação de sistemas de IA, bem como operadores, público consumidor e todas as pessoas impactadas pela utilização e resultados dessas aplicações.

Neste contexto, destaca-se a importância de identificar e categorizar os riscos envolvidos em todo o processo de produção, utilização e distribuição de ferramentas de IA. Embora nem sempre se enquadrem especificamente em uma categoria, devido à multiplicidade de formas de alcançar os resultados desejados, tanto em termos de programação de sistemas quanto de ferramentas e aplicações de diferentes formas de aprendizado de máquina ou outras abordagens técnicas, é fundamental analisar se os resultados esperados estão sendo alcançados, como estão sendo alcançados e quais são as implicações éticas desses resultados. Além disso, a localização geográfica e a conjunção de costumes de uma sociedade adequados em cada civilização também devem ser respeitadas e consideradas em cada estudo para manter IA ética.

Dentro dessa caracterização de possíveis perigos, foi pontuado com um pouco mais de destaque a abordagem dos “quatro perigos associados a falta de ética na IA”, a pirâmide de riscos elaborada pela UE para classificar em qual estágio cada sistema deve ser analisado, e a matriz de riscos que abrangem os potenciais perigos com a perda de controle. Essas três abordagens, quando combinadas de diversas maneiras, enriquecem a análise ética da IA, garantindo que um agente inteligente opere de maneira mais harmoniosa dentro da sociedade. Isso implica que futuras pesquisas serão necessárias para desbravar, adaptar e parametrizar outras abordagens, ou até mesmo combiná-las para garantir que a ética na IA seja assegurada em seus diversos contextos.

Foi possível destacar abordagens estudadas atualmente como mecanismos que buscam garantir que esses sistemas estejam alinhados com os valores éticos fundamentais, como a Auditoria Ética Baseada em Sistemas de Tomada de Decisão Automatizada. No

entanto, é importante ressaltar que ainda é necessária uma aplicação extensiva desse mecanismo regulatório no campo para que sua efetividade seja avaliada em diversos contextos, assim como as relações entre as organizações.

No que diz respeito à criação de agentes inteligentes, é crucial estar atento à quantidade e qualidade dos dados disponibilizados como referência em seus mecanismos. Além disso, torna-se necessário tornar claro para o usuário onde e como os dados que forneceram, seja por meio de permissões virtuais ou do próprio input de dados, estão sendo utilizados. Portanto, destaca-se a necessidade de modelos interpretáveis, em vez de explicáveis, para evitar uma “caixa preta” em relação aos dados e à ética dentro das operações dos conjuntos e IA.

Dessa forma, a ferramenta de Auditoria Ética - EBA ganha destaque como uma alternativa para os problemas éticos enfrentados na aplicação de IA. Reconhece-se a importância crítica de incorporar uma perspectiva ética desde as fases iniciais do desenvolvimento tecnológico.

Contudo, a maior parte das soluções investigada nesta pesquisa sobre inteligência artificial ainda se encontram em estágio experimental, em sua maioria, em investigações acadêmicas. Em paralelo, foi as implicações da regulamentação europeia de dados, bem como os exemplos concretos de sua aplicação, e citações sobre como anda o estudo e implementações de iniciativas que envolvem temática no contexto brasileiro, como destacado os mecanismos regulatórios por meio de leis como LGPD.

Assim, ao explorar os fundamentos éticos propostos na IA e na engenharia, bem como ao avaliar os impactos sociais da IA, uma proposta é seguir o caminho delineado por Floridi para uma abordagem ética e responsável em relação à tecnologia. Suas contribuições fornecem uma base sólida para o desenvolvimento futuro de políticas e práticas que garantam que a IA beneficie a sociedade como um todo, enquanto protege os valores e direitos fundamentais de todos os cidadãos. Percebe-se também que, em todos os contextos do desenvolvimento, sua pesquisa tem algo a acrescentar para agregar valor ao que vem sendo desenvolvido e ainda virá no futuro próximo.

No âmbito da engenharia, a integração da IA também se mostra cada vez mais vital. A colaboração entre engenheiros e especialistas em IA é essencial para garantir a criação de sistemas éticos e eficientes. Além disso, a aplicação de princípios éticos desde as fases conceituais dos projetos de engenharia é um componente crucial para o desenvolvimento responsável e sustentável das tecnologias. Observa-se uma das formas que isso pode ocorrer no desenvolvimento por meio do compromisso com o código de conduta e ética que todo profissional ao finalizar seus estudos declara, mas também destaca-se que devem ser exploradas outras maneiras de assegurar ações éticas que valorizem o impacto futuro da conduta na sociedade.

Fica claro que a rápida evolução da IA e sua integração na engenharia demanda uma abordagem ética robusta. A regulamentação europeia de dados oferece um modelo importante para orientar a implementação responsável de sistemas, servindo como referência para futuras discussões e legislações no Brasil e em outras partes do mundo.

Em suma, este estudo destaca a importância crítica de integrar considerações éticas na evolução da IA e engenharia. E assim, dar destaque o avanço tecnológico e garantimos que esse progresso seja sustentável, responsável e benéfico para a sociedade como um todo. O caminho adiante envolve a contínua colaboração entre a comunidade acadêmica, os reguladores e os praticantes da indústria, para criar um futuro digital ético e promissor.

Trabalhos Futuros

Considerando a complexidade e a constante evolução deste campo, torna-se necessário mais de uma recomendação para trabalhos futuros, que envolvam as mais diversas áreas de conhecimento.

Recomenda-se explorar mais aprofundadamente a aplicação prática da EBA em contextos específicos de IA e engenharia, analisando casos reais e os desafios por organizações ao incorporar essa abordagem. A realização de estudos de caso detalhados sobre a implementação de práticas éticas em projetos reais de IA e engenharia proporciona valiosos *insights*

Explorar e criar novos métodos e técnicas para avaliar e promover a ética na IA. A criação de *frameworks* e diretrizes específicas para diferentes tipos de sistemas de IA e aplicações de engenharia representa um passo importante.

Pode ser enriquecedor realizar outras comparações entre as mais diversas abordagens éticas que estão em desenvolvimento, ao mesmo tempo em que se investiga o estudo de outros filósofos e pesquisadores a respeito da IA e a visão desses sobre a correlação de ética e IA.

Explorar a interseção entre ética, IA e engenharia a partir de uma perspectiva multidisciplinar, envolvendo colaborações entre profissionais de diferentes campos, como filosofia, ética aplicada, ciência da computação e engenharia. Investigar a influência das diferentes culturas e contextos regionais nas considerações éticas na IA. Isso envolve a análise das regulamentações e políticas em diferentes países e regiões.

Realizar estudos sobre o impacto social de sistemas de IA em diferentes setores, como saúde, educação, transporte, entre outros, considerando as implicações éticas e sociais. Desenvolver métricas e métodos para avaliar de forma mais precisa e abrangente o impacto ético dos sistemas de IA, não apenas em termos de conformidade com regulamentações, mas também em relação aos valores e princípios fundamentais.

Além das abordagens mencionadas na pesquisa, o desenvolvimento de métricas e métodos para avaliar o impacto ético dos sistemas de IA, indo além da conformidade com regulamentações, é crucial para garantir a responsabilidade e a transparência. a criação de códigos de conduta e ética mais específicos para todos os profissionais envolvidos no desenvolvimento e implementação de sistemas de IA e engenharia é uma maneira tangível de garantir a integridade e a responsabilidade.

Essas sugestões representam diferentes direções que futuras pesquisas podem tomar para aprimorar a ética na interseção entre IA, engenharia e sociedade. Cada uma dessas

áreas oferece oportunidades valiosas para contribuir para o desenvolvimento de políticas e práticas que promovam a IA ética e responsável.

Referências

- GOOGLE Trends. Pesquisa sobre tendências de busca. Alphabet. Disponível em: <https://trends.google.com.br>. Acesso em: 25 ago. 2023. Citado 1 vez nas páginas 9, 24.
- ALVES, Marco Antônio Sousa; ANDRADE, Otávio Morato de. Da “caixa-preta” à “caixa de vidro”: o uso da explainable artificial intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. *Direito Público*, v. 18, n. 100, 2021. Citado 3 vezes nas páginas 15, 35, 41.
- CONHEÇA Alexa? Amazon. Disponível em: <https://www.amazon.com.br/b?ie=UTF8&node=19949683011>. Acesso em: 22 jul. 2023. Citado 1 vez na página 24.
- ASIMOV, Isaac. *Eu, robô*. Aleph, 2015. ISBN 978-85-7657-203-9. Citado 2 vezes nas páginas 13, 29.
- ACM Code of Ethics and Professional Conduct. Code of Ethics. Association for Computing Machinery. Jun. 2018. Disponível em: <https://www.acm.org/code-of-ethics>. Acesso em: 9 jul. 2023. Citado 2 vezes na página 43.
- BADUGE, Shanaka Kristombu et al. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, v. 141, p. 104440, 2022. ISSN 0926-5805. DOI: <https://doi.org/10.1016/j.autcon.2022.104440>. Citado 0 vez na página 22.
- BOSTROM, Nick; YUDKOWSKY, Eliezer. The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence*, Cambridge, v. 1, p. 316–334, 2014. Citado 3 vezes nas páginas 25, 35.
- CABRERA, Angel Alexander et al. FAIRVIS: Visual Analytics for Discovering Intersectional Bias in Machine Learning. In: 2019 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE, out. 2019. DOI: [10.1109/vast47406.2019.8986948](https://doi.org/10.1109/vast47406.2019.8986948). Disponível em: <https://doi.org/10.1109/vast47406.2019.8986948>. Citado 1 vez na página 40.
- CHATILA, Raja; HAVENS, John C. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Robotics and well-being*, 2019. DOI: [10.1007/978-3-030-12524-0_2](https://doi.org/10.1007/978-3-030-12524-0_2). Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-12524-0_2. Acesso em: 23 jul. 2023. Citado 1 vez na página 29.
- CONGRESSO NACIONAL DO BRASIL. *Projeto de Lei nº 2338*. 2023. Número 2338, CTIA - (Comissão Temporária Interna sobre Inteligência Artificial no Brasil). Disponível em: <https://bit.ly/3ES0Tmr>. Acesso em: 25 ago. 2023. Citado 1 vez na página 30.

- DE GÓMEZ, Maria Nelida Gonzalez. Luciano Floridi e os problemas filosóficos da informação: da representação à modelização. por. *InCID: Revista de Ciência da Informação e Documentação*, São Paulo, v. 4, n. 1, p. 3–35, 2013. Citado 1 vez na página 16.
- ESTEVA, Andre et al. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, Nature Publishing Group, v. 542, n. 7639, p. 115–118, 2017. Citado 1 vez na página 27.
- HIGH-LEVEL expert group on artificial intelligence. European Commission. Disponível em: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>. Acesso em: 1 set. 2023. Citado 1 vez na página 30.
- FLORIDI, Luciano. AI and its new winter: from myths to realities. eng. *Philosophy & technology*, Springer Netherlands, Dordrecht, v. 33, n. 1, p. 1–3, 2020. ISSN 2210-5433. Citado 3 vezes nas páginas 17, 21, 24.
- FLORIDI, Luciano. *Information: A very short introduction*. OUP Oxford, 2010. Citado 1 vez na página 17.
- FLORIDI, Luciano. *The fourth revolution: how the infosphere is reshaping human reality*. OUP Oxford, 2014. Citado 5 vezes nas páginas 18, 25, 26, 28.
- FLORIDI, Luciano. *The Onlife Manifesto*. Springer International Publishing, 2015. ISBN 3-319-04092-8. Citado 3 vezes nas páginas 18, 19, 42.
- FLORIDI, Luciano. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, 2019. DOI: [10.1007/s13347-019-00354-x](https://doi.org/10.1007/s13347-019-00354-x). Citado 6 vezes nas páginas 30, 31, 33, 34, 46.
- FLORIDI, Luciano et al. CapAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. *Social Science Research Network*, 2022. DOI: [10.2139/ssrn.4064091](https://doi.org/10.2139/ssrn.4064091). Citado 1 vez na página 45.
- FRISCHMANN, Brett; SELINGER, Evan. *Re-engineering humanity*. Cambridge University Press, 2018. Citado 1 vez na página 42.
- GALINARI RANGEL, et al. Comércio eletrônico, tecnologias móveis e mídias sociais no Brasil. *BNDES Setorial*, n. 41, p. 177–180, 2015. Citado 1 vez na página 19.
- GARDNER, Howard. *Inteligências Múltiplas: A Teoria na Prática*. Penso, 1995. ISBN 8573074132. Citado 1 vez na página 12.
- GERDES, J. Christian; THORNTON, Sarah M. Implementable Ethics for Autonomous Vehicles. In: *Autonomous Driving: Technical, Legal and Social Aspects*. Edição: Markus Maurer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. P. 87–102. ISBN 978-3-662-48847-8. DOI: [10.1007/978-3-662-48847-8_5](https://doi.org/10.1007/978-3-662-48847-8_5). Disponível em: https://doi.org/10.1007/978-3-662-48847-8_5. Acesso em: 25 set. 2023. Citado 1 vez na página 27.

- GLOCK, Rosana Soibelman; GOLDIM, José Roberto. Ética profissional é compromisso social. *Mundo Jovem. Porto Alegre: PUCRS*, p. 2–3, 2003. Citado 2 vezes na página 43.
- GULSHAN, Varun et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *jama*, American Medical Association, v. 316, n. 22, p. 2402–2410, 2016. Citado 1 vez na página 27.
- GUNNING, David. Broad agency announcement explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA), Tech. Rep.*, 2016. Citado 3 vezes nas páginas 36–38.
- IEEE Code of Ethics. The following is from the IEEE Policies, Section 7 - Professional Activities (Part A - IEEE Policies). Download a copy of the IEEE Code of Ethics I Adopted by the IEEE Board of Directors and incorporating revisions through June 2020. IEE Institute of Electrical e Electronics Engineers. Disponível em: <https://www.ieee.org/about/corporate/governance/p7-8.html>. Acesso em: 23 jul. 2023. Citado 1 vez na página 43.
- KRAFFT, Tobias D.; ZWEIG, Katharina A.; KÖNIG, Pascal D. How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance*, v. 16, n. 1, p. 119–136, 2022. DOI: <https://doi.org/10.1111/rego.12369>. Citado 3 vezes nas páginas 33, 34, 38.
- LIMA, Fabiano Guasti et al. Aplicação de redes neurais na análise e na concessão de crédito ao consumidor. *Revista de Administração-RAUSP*, Universidade de São Paulo, v. 44, n. 1, p. 34–45, 2009. Citado 1 vez na página 27.
- MADIEGA, Tambiama André. Artificial intelligence act. *European Parliament: European Parliamentary Research Service*, 2021. Citado 11 vezes nas páginas 30–34, 39, 40, 45.
- MARTINEZ, Luciano; MALTEZ, Mariana. O direito fundamental à proteção em face da automação. Universidade do Estado do Amazonas, 2019. Citado 1 vez na página 26.
- MOKANDER, Jakob et al. Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. *arXiv: Computers and Society*, 2021. DOI: [10.1007/s11948-021-00319-4](https://doi.org/10.1007/s11948-021-00319-4). Acesso em: 25 set. 2023. Citado 7 vezes nas páginas 27, 29, 30, 35, 39–41.
- MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos Sobre Aprendizado de Máquina. In: *SISTEMAS Inteligentes Fundamentos e Aplicações*. 1. ed. Barueri-SP: Manole Ltda, 2003. P. 89–114. Citado 1 vez na página 22.
- MUNN, Luke. The uselessness of AI ethics. *AI and Ethics*, Springer, v. 3, n. 3, p. 869–877, 2023. Citado 1 vez na página 46.
- NILSSON, Nils J. *Entre dados e robôs: Ética e privacidade na era da hiperconectividade*. Morgan Kaufmann, 1998. ISBN 1-55860-535-5. Citado 1 vez na página 15.

- NOVAES, Henrique Tahan; DAGNINO, Renato. O fetiche da tecnologia. *Org & Demo*, v. 5, n. 2, p. 189–210, 2004. Citado 3 vezes nas páginas 25, 26.
- CHATGPT. Resposta. OpenAI. Disponível em: <https://chat.openai.com/>. Acesso em: 23 jul. 2023. Citado 1 vez na página 24.
- PINHEIRO, Patricia Peck. *Proteção de dados pessoais: Comentários à lei n. 13.709/2018-lgpd*. Saraiva Educação SA, 2020. Citado 1 vez na página 29.
- PITEIRA, Martinha; APARICIO, Manuela; COSTA, Carlos J. A ética na inteligência artificial: Desafios. *A ética na inteligência artificial: Desafios*, IEEE, 2019. Citado 3 vezes nas páginas 43–45.
- RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 2019. DOI: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x). Citado 4 vezes nas páginas 35–37.
- RUSSEL, Stuart J. *Inteligência Artificial*. Elsevier Editora Ltda., 2013. ISBN 978-85-352-3701-6. Citado 7 vezes nas páginas 12, 13, 20, 27, 28, 44.
- SAKURAI, Ruudi; ZUCHI, Jederson Donizete. As revoluções industriais até a indústria 4.0. por. *Revista Interface Tecnológica*, v. 15, n. 2, p. 480–491, 2018. Citado 1 vez na página 20.
- SALEM, Fadi. A Smart City for public value: Digital transformation through agile governance—the case of Smart Dubai. *World government summit publications*, 2016. Citado 1 vez na página 40.
- SANTAELLA, Lucia; KAUFMAN, Dora. Os dados estão nos engolindo? eng ; por. *Civitas (Porto Alegre, Brazil)*, Pontifícia Universidade Católica do Rio Grande do Sul, v. 21, n. 2, p. 214–223, 2021. ISSN 1519-6089. Citado 3 vezes na página 42.
- SOPHOCLEOUS, Marios. Towards Fully Automated Decision-Making Systems for Precision Agriculture: Soil Sensing Technologies – “The Missing Link”. *null*, 2021. DOI: [10.4018/978-1-7998-5000-7.ch004](https://doi.org/10.4018/978-1-7998-5000-7.ch004). Citado 1 vez na página 27.
- TEFFÉ, Chiara Spadaccini de; MORAES, Maria Celina Bodin de. Redes sociais virtuais: privacidade e responsabilidade civil. Análise a partir do Marco Civil da Internet. *Pensar-Revista de Ciências Jurídicas*, v. 22, n. 1, p. 108–146, 2017. Citado 1 vez na página 28.
- TESLA Factory Tour with Elon Musk! Disponível em: https://youtu.be/mr9kK0_7x08?si=rCS01h7tyhVvV8ih. Acesso em: 25 ago. 2023. Citado 1 vez na página 23.
- UNICEF. *Declaração Universal dos Direitos Humanos (1948)*. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 29 ago. 2023. Citado 2 vezes nas páginas 14, 32.