

# MODELO DE GOVERNANÇA ÉTICA PARA IA EM ORGANIZAÇÕES: UMA ABORDAGEM VIA DESIGN SCIENCE RESEARCH PARA CONDUTAS CIDADÃS

ETHICAL GOVERNANCE MODEL FOR AI IN ORGANIZATIONS: A DESIGN  
SCIENCE RESEARCH APPROACH FOR CITIZEN CONDUCTS

Ciências Exatas e da Terra, Ciências Sociais Aplicadas • 13/06/2026

REGISTRO DOI: [10.70773/revistatopicos/781231563](https://doi.org/10.70773/revistatopicos/781231563)

Gilmar dos Santos Marques<sup>1</sup>

Lílian Santos Marques Severino<sup>2</sup>

Fábio Lúcio Lopes de Mendonça<sup>3</sup>

Edna Dias Canedo<sup>4</sup>

## RESUMO

Este artigo tem como objetivo identificar e analisar princípios éticos em Inteligência Artificial (IA) e propor um modelo inovador de governança, buscando orientar a adoção de práticas éticas em sistemas de IA em organizações públicas e privadas. Para tanto, nos procedimentos metodológicos foi empregado o Design Science Research (DSR) em sete fases, incluindo uma revisão bibliométrica realizada em 14 de abril de 2025, nas bases Scopus e Web of Science. A pesquisa utilizou as palavras-chave "*artificial intelligence*", "*governance*" e "*ethics*", resultando em uma amostra final de 76 artigos após a eliminação de duplicatas e o descarte de trabalhos não publicados em periódicos científicos. Os resultados culminaram na proposição de um Modelo de Governança Ética para IA em Organizações, composto por seis componentes interligados: Princípios Orientadores e Código de Ética de IA (POCEIA), Processo de Avaliação de Impacto Ético da IA e Gerenciamento de Riscos (AIEA), Estrutura de Auditoria Ética e Algorítmica (EAEA), Estrutura de Engajamento de Stakeholders e Feedback Contínuo (EESFC), Sistema de Monitoramento e Adequação Regulatória (SMAR) e Programa de Conscientização e Capacitação Ética em IA e Definição de Papéis (PCCEIA). A originalidade do trabalho reside em preencher a lacuna entre a teoria ética abstrata e a aplicação prática da IA nas organizações, transpondo o "que fazer" para o "como fazer" no desenvolvimento de uma IA transparente, justa e responsável. Como recomendação para estudos futuros, sugere-se a avaliação empírica do modelo em contextos organizacionais reais, conforme previsto na Fase 5 do DSR, para validar sua utilidade e eficácia.

**Palavras-chave:** Inteligência Artificial; Governança; Ética em IA; DSR.

## ABSTRACT

This article aims to identify and analyze ethical principles in Artificial

Intelligence (AI) and propose an innovative governance model, seeking to guide the adoption of ethical practices in AI systems in public and private organizations. To this end, the methodological procedures employed Design Science Research (DSR) in seven phases, including a bibliometric review conducted on April 14, 2025, in the Scopus and Web of Science databases. The research used the keywords "artificial intelligence," "governance," and "ethics," resulting in a final sample of 76 articles after eliminating duplicates and discarding works not published in scientific journals. The results culminated in the proposal of an Ethical Governance Model for AI in Organizations, composed of six interconnected components: Guiding Principles and AI Code of Ethics (POCEIA), AI Ethical Impact Assessment and Risk Management Process (AIEA), Ethical and Algorithmic Audit Framework (EAEA), Stakeholder Engagement and Continuous Feedback Framework (EESFC), Monitoring and Regulatory Compliance System (SMAR), and AI Ethical Awareness and Training Program and Role Definition (PCCEIA). The originality of the work lies in filling the gap between abstract ethical theory and the practical application of AI in organizations, translating the "what to do" into the "how to do" in the development of transparent, fair, and responsible AI. As a recommendation for future studies, empirical evaluation of the model in real organizational contexts is suggested, as foreseen in Phase 5 of the DSR, to validate its usefulness and effectiveness.

**Keywords:** Artificial Intelligence; Governance; Ethics in AI; DSR.

## 1. INTRODUÇÃO

A Inteligência Artificial (IA) tem se consolidado como uma das tecnologias mais transformadoras de nossa era, impulsionando inovações em praticamente todos os setores da sociedade e

redefinindo a forma como as pessoas interagem com o mundo digital e físico. A rápida evolução e a crescente integração de sistemas de IA, desde algoritmos de recomendação até ferramentas de tomada de decisão em áreas críticas como saúde e finanças, trouxeram à tona a urgência de debates profundos sobre suas implicações éticas e a necessidade de mecanismos robustos de governança.

À medida que a IA se torna mais autônoma e complexa, torna-se imperativo assegurar que seu desenvolvimento e aplicação estejam alinhados com valores humanos fundamentais e com o bem-estar social (Floridi, 2024a; Taddeo, 2025a). A discussão abrange desde a definição de princípios éticos basilares, como beneficência, não-maleficência, justiça, responsabilidade, equidade, transparência e privacidade (Floridi et al., 2018), até a complexa tarefa de traduzir esses conceitos abstratos em práticas concretas e verificáveis.

A operacionalização da ética em IA exige uma transição do "que fazer" para o "como fazer" (Wang, 2022a), demandando a criação de frameworks que permitam às organizações integrarem a ética em todo o ciclo de vida da tecnologia. Isso implica abordar questões éticas específicas, como o viés algorítmico, os "danos ocultos" e a erosão da autonomia humana, bem como a implementação de auditorias éticas e de mecanismos de prestação de contas, que garantam a construção de uma "IA confiável" (Stahl, 2024; Lütge, 2023c). A importância de uma "soft ethics", que complementa a legislação para enfrentar os desafios emergentes do ambiente digital, também tem sido enfatizada (Floridi, 2018a).

Nesse cenário, diversos autores, incluindo Pagallo (2021), têm contribuído para a formulação de arcabouços éticos que servem de

base para o desenvolvimento de IA mais consciente e responsável.

Diante do contexto apresentado, o problema central de pesquisa a ser respondido é: como as abordagens de ética e governança podem contribuir nas organizações gerando modelos de IA orientados por condutas cidadãs?

Para responder à pergunta-problema, este trabalho busca contribuir para a discussão crítica sobre o tema ao estabelecer como objetivo identificar e analisar os princípios éticos em IA, além de propor um modelo inovador de governança. Esse modelo tem como finalidade possibilitar a avaliação e a orientação para a adoção de práticas éticas em sistemas de IA, tanto em organizações públicas quanto privadas.

A proposta deste modelo visa preencher a lacuna entre a teoria e a prática na ética da IA, oferecendo um guia prático que ajudará as organizações a identificarem e corrigirem falhas em suas práticas atuais, promovendo um desenvolvimento de IA mais transparente, justo e, acima de tudo, responsável.

Este artigo foi estruturado em seis partes para facilitar a compreensão do tema e da metodologia adotada: 1) introdução, que contextualiza o problema e apresenta os objetivos do estudo; 2) revisão de literatura, que explora os padrões de abordagens sobre ética e governança em IA; 3) procedimentos metodológicos, detalhando a metodologia de Design Science Research utilizada; 4) resultados e discussões, onde se apresenta a análise dos dados; 5) modelo de governança ética para IA em organizações: uma abordagem via Design Science Research, que descreve o artefato

desenvolvido; e 6) conclusão, que sintetiza as principais descobertas e aponta para futuras pesquisas.

## **2. REVISÃO DA LITERATURA**

A revisão da literatura analisou os diferentes padrões observados nas abordagens de ética e governança em IA apresentadas a seguir.

### **2.1. Padrões de Abordagens Sobre Ética em IA**

Os artigos analisados revelam uma teia complexa de reflexões sobre ética em IA, que podem ser agrupadas nos seguintes padrões principais:

#### **2.1.1. Frameworks Éticos Gerais e Princípios Fundamentais**

A discussão sobre frameworks éticos gerais e princípios fundamentais constitui um padrão proeminente na literatura sobre ética em IA, com diversos autores propondo e explorando conjuntos de diretrizes que devem orientar o desenvolvimento e o uso da IA. Esses princípios incluem frequentemente beneficência, não-maleficência, autonomia, justiça, explicabilidade, responsabilidade, equidade, transparência e privacidade. Esses conceitos, quando aplicados corretamente, servem de base moral para assegurar que a IA contribua positivamente para a sociedade e respeite os valores humanos (Floridi, 2018a; Guimarães *et al.*, 2025; Taddeo, 2025a).

Floridi (2024a) destacou a importância de um pragmatismo sociotécnico na ética da IA ao sublinhar a necessidade de abordar os desafios éticos de forma prática. Além disso, detalhou os princípios-chave para uma IA benéfica e introduziu o conceito de "*soft ethics*", como complemento à legislação para enfrentar os desafios

emergentes do ambiente digital (Floridi, 2021a; Floridi et al., 2018; Floridi, 2018a).

Observa-se que o debate sobre ética em IA vem crescendo de forma acelerada e consistente. Surge uma visão abrangente sobre a integração do raciocínio ético na governança da IA, assim como a defesa da inclusão de considerações éticas desde as fases iniciais do desenvolvimento de modelos de IA (Taddeo, 2025a). As discussões começam a explorar a complexidade da governança global da IA, a necessidade de uma abordagem ética mais ampla, que, inclusive, requer valores normativos como equidade e privacidade, além de discutir a aplicação ética da IA nos Objetivos de Desenvolvimento Sustentável (ODS) (Taddeo, 2024a; Taddeo, 2023b).

Wang contribui para este campo ao propor diretrizes para a aplicação ética da IA na prática médica (Wang et al., 2023). Adicionalmente, em 2022, Wang apresenta um framework de quatro dimensões, abrangendo ética de dados, de algorítmica, de impacto social e de pesquisa, concebido para transformar princípios éticos em ações concretas (Wang, 2022a).

Os princípios do AI4People são amplamente detalhados e consolidam uma estrutura ética robusta (Lütge, 2021; Floridi et al., 2018). A abordagem da importância da IA confiável aprofunda a discussão sobre equidade e transparência na governança da IA. Também apresenta um complemento no framework ético e consolida a amplitude de abordagem nesse campo (Stahl, 2024; Roberts e Cowls, 2023; Pagallo, 2021).

### **2.1.2. Questões Éticas Específicas e Desafios Contextuais**

A literatura sobre ética em IA dedica atenção significativa a problemas éticos que emergem em domínios específicos de sua aplicação, transcendendo as discussões de princípios gerais para abordar dilemas práticos. Esses problemas incluem preconceito algorítmico, o que pode levar a resultados discriminatórios; a ocorrência de "danos ocultos", ou seja, impactos negativos difíceis de detectar e quantificar (Floridi, 2024c); questões de privacidade e a necessidade de proteger dados sensíveis; os efeitos da IA no bem-estar digital dos usuários (Floridi, 2020a); a manipulação e erosão da autonomia humana (Taddeo, 2023c); preocupações com justiça social, vigilância e os complexos desafios éticos que surgem na interação entre humanos e sistemas de IA ou entre humanos e máquinas.

Nesse contexto, o tratamento dos "danos ocultos" (hidden harms) da IA, que são difíceis de mensurar, requer a necessidade de uma abordagem mais abrangente para a garantia de um modelo de IA responsável, que seja capaz de atender as implicações éticas das cidades inteligentes, por meio da ética dos algoritmos, identificando problemas e soluções-chave, para as questões éticas relacionadas ao bem-estar digital (Floridi, 2024c; Floridi, 2024d; Floridi, 2022a; Floridi, 2020a).

Taddeo (2024b) contribui com uma taxonomia ética para a Inteligência de Código Aberto (OSINT), abordando privacidade e manipulação de dados, e em 2023, discute como a IA pode influenciar a agência humana. Além de examinar questões de compartilhamento de dados, também propôs uma taxonomia para a ética dos algoritmos, categorizando as dimensões éticas intrínsecas, extrínsecas e sistêmicas (Taddeo, 2022b).

Wang examinou o impacto da IA na medicina chinesa, abordando desafios éticos e de governança, e analisou a relação entre IA e sustentabilidade, incluindo os riscos associados à proteção de dados (Wang, 2024a; Wang, 2023a). Lütge analisou a complexa questão da responsabilidade na interação humano-IA e, em outro trabalho, aborda os desafios éticos da governança de IA durante a pandemia de COVID-19 (Lütge, 2023b). Complementarmente, Pagallo discute barreiras éticas e legais no uso da IA no setor de saúde, incluindo questões de privacidade, preconceito e transparência, e explora a "caixa preta" da comunicação humano-IA, propondo uma agenda regulatória para sua transparência (Pagallo, 2024a; Pagallo, 2024b).

### **2.1.3. Auditoria Ética e Operacionalização**

A transposição de princípios éticos abstratos para práticas concretas e verificáveis configura uma etapa crucial no desenvolvimento responsável da IA. Nesse panorama, surgem discussões sobre a implementação de auditorias éticas e sobre como conduzir avaliações de impacto moral e ético. Discute-se ainda o planejamento de sistemas sensíveis a valores, bem como o estabelecimento de mecanismos robustos para garantir uma "IA confiável". O propósito dessas iniciativas é impedir que as diretrizes éticas fiquem confinadas ao discurso teórico, promovendo sua integração prática ao longo de todo o ciclo de vida da IA.

Floridi é considerado um autor central nesta área, cujos estudos enfatizam fortemente que as auditorias baseadas em ética são imprescindíveis para o avanço da IA confiável. Em 2021, publicou trabalhos detalhando como tais auditorias podem ser elaboradas e aplicadas para avaliar a conformidade de sistemas de IA com os princípios éticos (Floridi, 2021b; Floridi, 2021c). Em uma publicação de

2020, o autor já havia explorado as ferramentas e os métodos disponíveis para traduzir os princípios éticos da IA em práticas acionáveis (Floridi, 2020b).

Outros estudiosos também têm contribuído consideravelmente para o desenvolvimento de metodologias de auditoria ética e de gerenciamento de riscos algorítmicos. Assim, em 2025, em uma das publicações, Taddeo propõe a integração do raciocínio ético na governança da IA. Em outra publicação, de 2023, discute a utilização estratégica da IA para propiciar ecossistemas digitais confiáveis, insinuando, ainda que de forma velada, a necessidade de mecanismos de verificação mais robustos (Taddeo e Floridi, 2023).

Mokander (2021a, 2021b) reúne um leque de investigações que, em detalhes minuciosos, traçam o contorno de métodos de auditoria ética, concentrando-se na contenção dos riscos algorítmicos e na sua aplicação prática em contextos científicos e de engenharia.

Lütge (2023c) estuda a transposição de princípios de IA confiável para a prática concreta, oferecendo um framework para o desenvolvimento e a implantação da IA, apresentando um modelo que funciona como bússola tanto para a concepção quanto para a adoção responsável da tecnologia.

Wang (2022a) examina a transição dos princípios éticos de alto nível para a sua aplicação prática, fornecendo um arcabouço ético para a investigação em ciências sociais sobre IA. Todos esses estudos, em conjunto ou isoladamente, contribuíram para a efetiva aplicação prática dos princípios de ética. Essas contribuições ressaltam a importância de ferramentas e processos concretos para

operacionalizar a ética na IA e garantem que os sistemas sejam planejados e aplicados de maneira responsável e confiável.

#### **2.1.4. Debates e Narrativas Éticas**

Na prática, a forma como a IA é percebida e, por consequência, o modo como a ética se desenvolve não surgem num vazio. A IA é continuamente esculpida pelas narrativas e pelos discursos predominantes na sociedade, que apontam para a existência de narrativas concorrentes exercendo forte influência sobre o campo da ética em IA, distinguindo entre abordagens que vão do “*ethics washing*” - uma adoção superficial de princípios éticos -, até a perspectiva de “*ethics as regulation*”, que busca formalizar a conduta da IA por meio de normas (Floridi, 2024a; Mokander, 2024).

Assim, observa-se que a literatura aponta para a necessidade de desenvolver abordagens mais amplas e inclusivas para a ética em IA sugerindo que uma compreensão completa e eficaz exige a integração de diversas perspectivas, superando as dicotomias e polarizações existentes (Taddeo, 2024c; Roberts, 2024).

### **2.2. Padrões de Abordagens Sobre Governança em IA**

A governança da IA é um tema igualmente complexo e multifacetado nos artigos, com padrões que incluem:

#### **2.2.1. Frameworks e Modelos de Governança Gerais**

Com a IA penetrando quase todas as esferas da vida contemporânea, sua regulação se consolidou como um campo de pesquisa imprescindível. Isso inclui a concepção de marcos regulatórios e de políticas públicas, além da implantação de

estratégias que guiem o uso da IA em contextos tão variados quanto os que a cercam.

Essa vertente da governança tem por meta conter os riscos, ampliar os benefícios, alinhar a tecnologia aos valores sociais e garantir que a implantação da IA ocorra de forma responsável. Cabe destacar que a investigação sobre bem-estar digital ressalta a cooperação entre múltiplas partes e reforça a urgência de adotar uma perspectiva holística na condução da tecnologia (Taddeo e Floridi, 2020; Monteiro & Ayrosa, 2025).

Stahl (2024), por sua vez, destaca-se como um dos mais veementes defensores da boa governança e da inovação responsável, insistindo que uma abordagem holística, acompanhada de supervisão regulatória rigorosa, é indispensável. Ele postula que a boa governança se configura como uma resposta imprescindível aos perigos que emergem da IA e, ao mesmo tempo, demonstra como a própria IA pode ser acionada para reforçar essa arquitetura regulatória (Stahl, 2021a; Stahl, 2021b). Ainda, aprofunda o conceito de inovação responsável, delineando possíveis trilhas para um desenvolvimento tecnológico que incorpore, de forma intrínseca, valores éticos (Stahl, 2022a).

Lütge (2023c) volta o olhar para a árdua missão de transformar princípios abstratos em práticas tangíveis, erigindo estruturas que pretendem orientar tanto o desenvolvimento quanto a implantação responsáveis da IA. Os princípios descritos na literatura, ao longo dos anos, conquistaram ampla adesão e hoje sustentam a base da governança participativa da IA (Lütge, 2021; Floridi et al., 2018). A perspectiva ganha profundidade graças às contribuições de Roberts e Cowls (2023), que oferecem um olhar interdisciplinar sobre os

modelos de governança, sublinhando a complexidade do tema e a necessidade de empregar múltiplas lentes para compreender e administrar a IA de maneira eficaz.

### **2.2.2. Governança e Políticas Regionais/nacionais**

As estratégias de governança da Inteligência Artificial (IA) variam significativamente entre regiões e países, refletindo diferentes prioridades culturais, econômicas e políticas. Esta abordagem concentra-se no exame desses marcos regulatórios específicos, com foco notável nas iniciativas da China, da União Europeia (UE) e do Reino Unido.

Floridi (2021d, 2023a, 2023b) tem se dedicado a comparações meticulosas das estratégias e dos marcos regulatórios de IA nesses três blocos, ampliando ainda a investigação da postura chinesa de 2021. Em seus trabalhos, ele evidencia as sutilezas e os pontos em que as políticas globais de IA convergem ou divergem. Wang (2024b) analisa as ramificações da Lei de IA da UE, centrando-se no setor de serviços financeiros e enfatizando a necessidade premente de auditorias algorítmicas formalizadas. Em paralelo, o autor examina a forma como a IA vem sendo disciplinada na medicina e esboça o panorama das políticas gerais chinesas, trazendo à baila diferentes estudos (Wang, 2024a; Wang, 2022b; Wang, 2021a).

Essa constelação de análises abre uma janela para observar a trajetória da Ásia que impulsiona o debate regulatório ao desmembrar as barreiras jurídicas, as quais ainda permeiam o setor de saúde, ao mesmo tempo em que propõe um conjunto de estratégias para tornar a comunicação entre humanos e IA mais transparente (Pagallo, 2024a; Pagallo, 2024b). A partir de 2017, ele

ainda revisita as repercussões que a inteligência artificial tem provocado nos sistemas jurídicos (Pagallo, 2017). Perceber essas diferentes abordagens regionais mostra-se crucial para construir políticas globais de IA que equilibrem eficácia e robustez ética.

### **2.2.3. Governança internacional/global**

A governança global da IA é reconhecida como uma área crítica que enfatiza a necessidade premente de coordenação e cooperação para equilibrar inovação tecnológica, considerações éticas e segurança. As complexidades inerentes à natureza transfronteiriça da IA exigem abordagens que transcendam as fronteiras nacionais e regionais.

Taddeo (2024a) aprofunda seus estudos sobre as complexidades da natureza da IA e sustenta que a governança global da IA só será eficaz se adotar um modelo híbrido que reúna direito internacional, padrões técnicos e diretrizes éticas. Essa lente panorâmica revela-se indispensável na concepção de arcabouços capazes de acomodar a pluralidade dos contextos culturais e regulatórios. Roberts (2024) amplia o debate sugerindo que é imprescindível uma conexão entre a esfera internacional e a ética em IA. Reconhece, ainda, que a geopolítica e as normas éticas precisam convergir para que a governança se torne realmente eficaz.

Fazer com que as políticas internas se encaixem nas estruturas globais é uma questão decisiva. Floridi (2023a) chama a atenção para isso ao comparar as estratégias de governança de IA adotadas na China e na União Europeia; já Wang (2021) analisa as diretrizes éticas chinesas sob o prisma das regras da OMC. Esses trabalhos revelam tanto os entraves quanto as possibilidades de sincronizar as

iniciativas nacionais com os objetivos de uma governança global mais coerente.

Já Lütge (2022a) mergulha na forma como a Europa aborda a governança da IA, analisando as diversas iniciativas da UE e destacando a centralidade dos valores humanos essenciais. Dando sequência ao debate, Stahl (2022b) propõe a criação de uma Agência Europeia de Inteligência Artificial, com o objetivo de tornar mais eficazes no continente a coordenação e a supervisão da IA.

#### **2.2.4. Governança Corporativa e Organizacional**

Com um olhar voltado para o futuro, Pagallo (2024c) destaca a necessidade de cooperação internacional para estabelecer novos marcos regulatórios no espaço exterior – um ambiente em que a IA já atua e cujo papel só tende a se aprofundar.

Dessa forma, a governança da IA nas corporações surge como um paradigma singular e de grande importância, orientando práticas e estratégias internas, buscando conduzir os sistemas de IA de forma ética e responsável. Numa perspectiva como esta, a análise de Floridi (2022b) sobressai ao desbravar tanto as armadilhas quanto as práticas recomendadas na governança corporativa da inteligência artificial, empregando o segmento biofarmacêutico como exemplo concreto.

#### **2.2.5. Auditoria Como Mecanismo de Governança**

Os resultados da sua pesquisa revelam o labirinto que se abre ao tentar alinhar a inovação tecnológica com dilemas éticos, regulatórios e de privacidade de dados, realçando a necessidade premente de criar comitês de ética em IA, de erigir estruturas

robustas de governança de dados e de fomentar uma colaboração interdisciplinar, de modo que as organizações possam abraçar a IA de forma proativa e responsável. Ele aprofunda a análise da natureza e dos limites das auditorias éticas de sistemas automatizados de decisão (Floridi, 2021b).

Em 2020, revisitou o conjunto de instrumentos disponíveis para traduzir princípios éticos em práticas concretas (Floridi, 2020b). Mais recentemente, Floridi (2024c) propõe uma perspectiva mais abrangente para a garantia da IA, incorporando a detecção de “danos ocultos”.

A metodologia e a operacionalização das auditorias éticas têm atraído o olhar de outros estudiosos. Taddeo e Floridi (2021) descrevem a auditoria ética de algoritmos como um novo mecanismo de governança capaz de possibilitar uma IA confiável. Já, Mokander (2021a, 2021b) descreve detalhadamente como auditorias ancoradas em princípios éticos podem amortecer os riscos algorítmicos, apontando, ainda, suas possíveis aplicações na ciência e na engenharia. A importância tanto das auditorias baseadas em princípios éticos quanto das auditorias algorítmicas se mostra essencial para uma governança de IA verdadeiramente eficaz, atuando como instrumentos indispensáveis para garantir responsabilidade, transparência e confiança nos sistemas de IA.

### **2.2.6. Governança em Contextos Específicos**

Ao analisar a governança da IA, revela-se sua natureza multifacetada, exigindo um olhar cuidadoso tanto para os obstáculos quanto para as oportunidades que surgem quando a tecnologia é inserida em ambientes particulares — sejam cidades

inteligentes, sistemas de saúde, iniciativas de sustentabilidade ou cenários de crise. Essa especificidade contextual, por sua vez, sugere a construção de marcos e políticas que se ajustem às particularidades e aos dilemas éticos de cada setor.

Com o aprofundamento do debate sobre a governança da IA nas cidades inteligentes, investiga-se as complexas questões éticas que surgem, como a IA pode impulsionar os ODS, ressaltando desafios e oportunidades que se repetem (Floridi, 2024d; Floridi, 2023c). No cenário corporativo investiga-se os obstáculos e as práticas de excelência na governança da IA nas organizações, com ênfase particular na indústria biofarmacêutica (Floridi, 2022b).

De forma complementar, Wang (2024a) tem direcionado seu olhar para a aplicação da IA em setores críticos, como as nuances éticas e os dilemas de governança da IA na medicina chinesa, analisando tanto os avanços quanto as dificuldades, e, também, avalia a Lei de IA da União Europeia, concentrando-se nas suas implicações para auditorias algorítmicas no contexto dos serviços financeiros (Wang, 2024b).

Em outro estudo, Wang (2023a) apresenta um quadro amplo e minucioso, contendo as oportunidades e os obstáculos que a IA traz à sustentabilidade, e ainda aparece como coautora das diretrizes para o uso ético da IA na prática clínica (Wang et al., 2023). Por fim, Lütge (2023b) analisa a governança da IA durante a pandemia de COVID-19, mostrando lições essenciais sobre como a tecnologia foi implementada e gerida num cenário de emergência global, enfatizando a necessidade de resiliência ética.

### **3. PROCEDIMENTOS METODOLÓGICOS**

A abordagem metodológica deste estudo foi delineada em sete etapas principais, conforme apresentado nas Figura 1 e 2. Inicialmente, realizou-se a seleção das bases de dados, optando-se pela Scopus e pela Web of Science (WoS). Estas bases foram escolhidas por reunirem artigos publicados em periódicos de reconhecido prestígio internacional e por possuírem elevado fator de impacto.

A busca foi realizada nas duas bases em 14 de abril de 2025. Para identificar publicações relevantes ao tema, utilizaram-se as palavras-chave: *("artificial intelligence" OR AI) AND governance AND ethics*. Na Scopus, os campos de busca considerados foram o título, o resumo e as palavras-chave dos artigos, resultando em 983 publicações retornadas.

**Figura 1.** Modelo de pesquisa na base de dados Scopus



**Fonte:** Scopus, a partir de pesquisa dos autores.

Na *Web of Science* (WoS), a pesquisa foi conduzida no campo *"topic"*, abrangendo título, resumo e palavras-chave (do autor e plus), com um total de 626 artigos identificados.

**Figura 2.** Modelo de pesquisa na base de dados Web of Science

**626** results from Web of Science Core Collection for:

("artificial intelligence" OR ai) AND governance AND ethics (Topic)



**Fonte:** Web of Science, a partir de pesquisa dos autores.

Para o tratamento e análise dos resultados, foram empregados os softwares R e RStudio, em conjunto com os pacotes Bibliometrix e Shiny. O Bibliometrix foi utilizado para importar e manipular os metadados, permitindo a aplicação de filtros e a organização dos dados. O pacote Shiny, por sua vez, proporcionou uma interface gráfica interativa, facilitando a visualização e análise das informações processadas.

Após a importação dos metadados das duas bases, os resultados foram integrados (merge) e submetidos a uma análise conjunta, cujo objetivo principal foi a eliminação de publicações duplicadas. Ao término desse processo, apresentou-se um total de 1.591 artigos. Após a exclusão das duplicidades, a amostra foi reduzida para 1.097 artigos.

Do total de 1.097 publicações resultantes do estudo bibliométrico, selecionaram-se 91 artigos mais citados para análise aprofundada. Destes 91 artigos, 15 foram descartados por corresponderem a trabalhos apresentados em eventos acadêmico-científicos, mas não publicados em periódicos científicos. Dessa forma, 76 artigos foram selecionados como objeto deste estudo. A análise dos resumos desses 76 artigos buscou identificar diferentes padrões e abordagens sobre ética e governança em IA, bem como o papel desses elementos na promoção de modelos de IA orientados por condutas cidadãs nas organizações, de acordo com os procedimentos de métodos e técnicas de pesquisa (Sordi; Meireles; Sanches, 2011).

## 4. RESULTADOS E DISCUSSÕES

As abordagens sobre ética e governança da IA, conforme detalhadas na literatura, delineiam um roteiro robusto para organizações que aspiram a desenvolver e implementar modelos de IA orientados por "condutas cidadãs". Essa orientação transcende a mera conformidade legal, buscando ativamente a criação de sistemas de IA que promovam o bem-estar social, respeitem os direitos individuais e operem de forma justa e transparente. A contribuição da pesquisa para este paradigma manifesta-se em diversas frentes interligadas:

Em primeiro lugar, é indispensável garantir que valores fundamentais estejam presentes já na fase de concepção. Ao adotar quadros éticos amplos e princípios basilares, como os propostos por Floridi (2024a, 2021a), Taddeo (2025a, 2023a), Wang (2023a), Lütge (2021) e Stahl (2024), as organizações conseguem incorporar, ainda nas etapas iniciais de design e desenvolvimento de sistemas de IA, valores como justiça, equidade, transparência, privacidade, responsabilidade e beneficência.

Em segundo lugar, ao realizar um exame minucioso das questões éticas que emergem em contextos particulares, consegue-se aprimorar a mitigação de riscos e de danos específicos. As contribuições de Floridi (2024c, 2024c, 2022b, 2020a), Taddeo (2024a, 2023b, 2022a, 2022b), Wang (2024b, 2023a), Lütge (2023a, 2023b) e Pagallo (2024c, 2024b) capacitam as organizações a identificarem, de forma proativa, desafios como o preconceito algorítmico, os "danos ocultos" (Floridi, 2024c), as violações de privacidade (Taddeo, 2022c) e o impacto na autonomia humana (Taddeo, 2023c).

Seguindo esse caminho, as organizações conseguem construir sistemas mais seguros, mais equitativos e menos propensos a causar danos inadvertidos aos cidadãos. Por exemplo, ao conceber IA para cidades inteligentes, incorporar, desde as fases iniciais, as preocupações de vigilância e justiça social (Floridi, 2024d), em sintonia com o debate atual, garante que a tecnologia beneficie todos os residentes, e não apenas um grupo seletivo.

Em terceiro lugar, a elaboração de mecanismos de prestação de contas e de transparência ganha força quando a auditoria ética passa a ser o ponto central da sua operacionalização. As contribuições de Floridi (2024c, 2021b, 2021c, 2020a), Taddeo (2021) e Mokander (2021a, 2021b) oferecem ferramentas práticas para monitorar, avaliar e validar o desempenho ético dos sistemas de IA. Entre elas, destacam-se as auditorias regulares que buscam detectar vieses, assegurar a explicabilidade e a transparência (Pagallo, 2024b) e garantir que as decisões da IA possam ser compreendidas e questionadas.

Em quarto lugar, vale destacar que o engajamento de múltiplos stakeholders é um componente vital. Os padrões de governança — sejam eles gerais (Taddeo, 2023a), regionais/nacionais (Floridi, 2023a; Wang, 2024b), internacionais (Taddeo, 2024a; Roberts, 2024) ou específicos a determinados contextos (Floridi, 2024d, 2023c; Lütge, 2023b) — enfatizam repetidamente a importância de uma colaboração estreita entre governos, sociedade civil, academia e o público. Lütge (2021) destaca a importância da governança participativa, e Taddeo (2023a) propõe modelos abrangentes que combinam cooperação multissetorial, assegurando que as preocupações e expectativas dos cidadãos sejam ouvidas e incorporadas ao longo do ciclo de vida da IA. Organizações que

adotam uma postura de “conduta cidadã” envolvem ativamente as comunidades para as quais a IA foi desenvolvida, buscam feedback e cocriam soluções que realmente atendam às suas necessidades e respeitem seus valores.

Em quinto lugar, adaptar-se aos marcos regulatórios e legais é imprescindível para que a IA exerça uma conduta verdadeiramente cidadã. Isso exige captar as sutilezas da governança e das políticas regionais ou nacionais, como nas análises de Floridi (2023a, 2023b) sobre a China, a UE e o Reino Unido; nos estudos de Wang (2024a, 2024b, 2022b, 2021a) focados na China; e na investigação de Lütge (2022a) acerca da Lei de IA da UE. Segundo a perspectiva europeia, e de acordo com Pagallo (2024a, 2017) ao tratar das barreiras jurídicas, as organizações não se restringem a observar apenas as normas em vigor – como o GDPR e a Lei da IA da UE –, mas também se preparam para as regulações que ainda estão por surgir.

Essa proatividade revela-se imprescindível para desenvolver IA com “conduta cidadã”, isto é, sistemas juridicamente robustos e socialmente aceitos, capazes de driblar armadilhas jurídicas e de forjar uma reputação de responsabilidade. A análise de Wang (2024b) sobre a Lei da IA e suas implicações para auditorias algorítmicas evidencia a necessidade de as organizações se alinhar a essas diretrizes.

Floridi (2018a) introduziu o termo “*soft ethics*” como complemento à legislação; Stahl (2021a, 2021b) tem se destacado como defensor da boa governança e da inovação responsável, ao passo que Taddeo (2025a) propõe inserir o raciocínio ético diretamente na estrutura de governança. Organizações que cultivam essa cultura dão aos seus colaboradores — de engenheiros a gestores de produto — a

liberdade e o estímulo para fazer escolhas éticas no cotidiano, questionando o status quo e perseguindo soluções que coloquem o bem público em primeiro plano. Com isso, estabelece-se um ambiente em que a “conduta cidadã” se incorpora ao DNA da organização e se espelha nos seus produtos de IA.

Em síntese, as correntes contemporâneas de ética e governança da IA disponibilizam tanto o alicerce filosófico quanto as ferramentas operacionais e os guias estratégicos de que as organizações precisam. Elas dão às empresas a possibilidade de abandonar um modelo de desenvolvimento de IA meramente técnico, para que possam aceitar, de forma consciente, as responsabilidades sociais e cívicas. Dessa maneira, os modelos de IA produzidos não se limitam a avanços tecnológicos; eles também se firmam em princípios de justiça, equidade, transparência e respeito à dignidade humana, verdadeiramente guiados por condutas cidadãs.

## **5. MODELO DE GOVERNANÇA ÉTICA PARA IA EM ORGANIZAÇÕES: UMA ABORDAGEM VIA DESIGN SCIENCE RESEARCH**

Neste estudo, apresenta-se um modelo de governança desejado para administrar os princípios éticos que regem a IA nas organizações, apoiado na abordagem Design Science Research (DSR). A DSR, ao articular pesquisa e design, cria saberes inovadores por meio do desenvolvimento de artefatos, com a finalidade de estimular um desenvolvimento de IA mais transparente, equitativo e responsável (Hevner et al., 2004). O artefato apresentado aqui é um modelo de governança que sistematiza as diferentes abordagens sobre ética e governança da IA, alinhado à literatura mais recente e às discussões da seção 4. O objetivo desse modelo é capacitar as

organizações a adotarem uma “conduta cidadã” em suas iniciativas de IA, ultrapassando a mera conformidade legal e buscando gerar um impacto social positivo.

## **5.1. Fases do Design Science Research e o Desenvolvimento do Artefato**

O processo de DSR tipicamente envolve as seguintes fases, que são aplicadas no desenvolvimento deste modelo de governança:

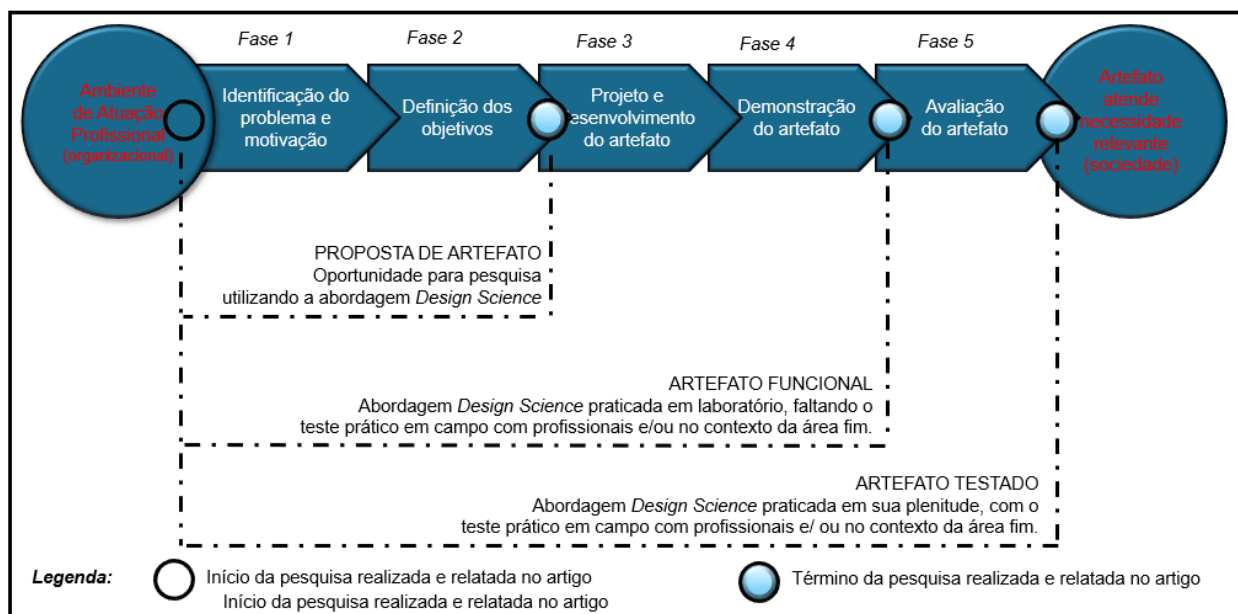
- **Fase 1** - Identificação do problema e motivação: o problema central é a lacuna entre os princípios éticos abstratos da IA e sua aplicação prática e sistemática nas organizações, podendo acarretar o risco de IAs não transparentes, injustas ou irresponsáveis. A motivação é a necessidade de um roteiro prático para que as organizações possam desenvolver IA com "conduta cidadã".
- **Fase 2** - Definição dos objetivos para uma solução: busca-se criar um modelo de governança que incorpore os valores centrais da IA ética, atenuar riscos específicos, estabeleça mecanismos de prestação de contas, estimule o engajamento dos stakeholders, assegure a conformidade regulatória e cultive uma cultura organizacional ética.
- **Fase 3** - Projeto e desenvolvimento do artefato (O Modelo de Governança): esta fase, o cerne do DSR, envolve a criação do próprio modelo de governança, articulando seus componentes e processos com base na literatura revisada.
- **Fase 4** - Demonstração do artefato: ilustrar a aplicabilidade do modelo em cenários organizacionais típicos de

desenvolvimento e implementação de IA. (Apresentado na seção 5.2).

- **Fase 5 - Avaliação do artefato:** avaliar a utilidade e eficácia do modelo em contextos reais. (Sugere-se pesquisas futuras.)

E, por último, tem-se o processo de comunicação dos resultados, que se faz por meio do compartilhamento do artefato e insights obtidos, conforme a figura 3, a seguir:

**Figura 3.** Modelagem do processo de DSR para o desenvolvimento do modelo de governança



**Fonte:** elaborado pelos autores, adaptado de Sordi *et al.* (2015)

## 5.2. O Modelo de Governança Ética de IA para Organizações (o Artefato)

O modelo de governança proposto é composto por seis componentes interligados, cada um projetado para atender a uma dimensão crítica da governança ética da IA, conforme os pilares identificados nos resultados e discussões.

### **5.2.1. Componente 1: Princípios Orientadores e Código de Ética de IA (POCEIA)**

Este componente estabelece a base moral e os valores que devem guiar todas as atividades relacionadas à IA na organização.

- **Descrição:** descreve um conjunto claro de princípios éticos — justiça, equidade, transparência, privacidade, responsabilidade e beneficência — juntamente com um código de conduta detalhado que serve de guia ao desenvolvimento e à aplicação da IA. Também traz diretrizes para colocar em prática o “design sensível a valores”.
- **Fundamentação teórica:** os resultados e as discussões deixam claro que inserir valores fundamentais já na fase de concepção é imprescindível. Floridi (2024a, 2021a), Taddeo (2025a, 2023a), Wang (2023a), Lütge (2021) e Stahl (2024) são autores de destaque que defendem a incorporação desses valores nas primeiras etapas do design de sistemas de IA, prevenindo ajustes posteriores e assegurando que a “conduta cidadã” se torne um pilar da arquitetura do sistema. Os princípios de equidade e justiça, como apontam Floridi (2018a) e Taddeo (2023a), revelam claramente que os algoritmos precisam ser imparciais.

### **5.2.2. Componente 2: Processo de Avaliação de Impacto Ético da IA (AIEA) E Gerenciamento de Riscos**

Este componente foca na identificação proativa e mitigação de riscos éticos específicos.

- Descrição: trata-se de um procedimento cuidadoso que procura identificar, examinar e abranger os riscos e os impactos éticos negativos que acompanham cada projeto de IA, desde a ideia inicial até sua implementação. Envolve a investigação de potenciais vieses algorítmicos, a revelação de danos ocultos e a análise de como tudo isso interfere na autonomia humana e na privacidade.
- Fundamentação teórica: ao aprofundar as análises das problemáticas éticas nos diferentes domínios, a mitigação de riscos e de danos pontuais ganha um impulso considerável (Floridi, 2024c, 2024d, 2022b, 2020a; Taddeo, 2024a, 2023b, 2022a, 2022b; Wang, 2024a, 2023a; Lütge, 2023a, 2023b; Pagallo, 2024a, 2024b). Floridi (2024c) sublinha a existência de “danos ocultos”, que precisam ser detectados de forma proativa. Quando se cruza a ótica de Taddeo (2023c, 2022a) acerca da privacidade e da autonomia humana com a inquietude de Floridi (2024c) sobre vigilância e justiça social nas cidades inteligentes, nasce o alicerce que sustenta a inserção dessas avaliações.

### **5.2.3. Componente 3: Estrutura de Auditoria Ética e Algorítmica (EAEA)**

Este componente garante a responsabilidade e transparência dos sistemas de IA.

- Descrição: um conjunto de procedimentos e ferramentas projetados para conduzir auditorias éticas – internas ou externas – em sistemas de IA. Prioriza a clareza das decisões da IA, a transparência dos processos subjacentes, a detecção e a

correção de vieses, além da rastreabilidade necessária para garantir prestação de contas. Também abrange os requisitos de documentação e a elaboração dos relatórios de auditoria.

- **Fundamentação Teórica:** a construção de mecanismos que garantam a prestação de contas e a transparência ganha força ao se colocar ênfase na auditoria ética e na operacionalização. Floridi (2024c, 2021b, 2021c, 2020a) destaca-se como um dos principais defensores da auditoria ética como caminho para uma IA confiável. Taddeo e Floridi (2021) apresentam a auditoria ética de algoritmos como um recurso de governança, enquanto Mokander (2021a, 2021b) elabora metodologias voltadas à mitigação de riscos algorítmicos. A formalização das auditorias algorítmicas, examinada por Wang (2024b) no âmbito da Lei de IA da UE, realça a relevância cada vez maior dessas ferramentas no panorama regulatório. Pagallo (2024b) também enfatiza a necessidade de assegurar explicabilidade e transparência.

#### **5.2.4. Componente 4: Estrutura de Engajamento de Stakeholders e Feedback Contínuo (EESFC)**

Este componente assegura que a IA seja desenvolvida com a contribuição e em benefício das comunidades afetadas.

- **Descrição:** cria, de forma organizada, canais e procedimentos formais que possibilitam a consulta e a cooperação tanto com colaboradores internos quanto com agentes externos — clientes, representantes da sociedade civil, universidades e entidades governamentais. Incentiva a escuta cuidadosa das inquietudes e aspirações da população, garantindo que o

feedback seja incorporado ao longo de todo o ciclo de vida da IA.

- **Fundamentação teórica:** a colaboração de diversos stakeholders mostra-se indispensável às normas de governança, sejam elas de âmbito geral (Taddeo, 2023a), de alcance regional ou nacional (Floridi, 2023a; Wang, 2024b), de esfera internacional (Taddeo, 2024a; Roberts, 2024) ou ainda em contextos específicos (Floridi, 2024d, 2023c; Lütge, 2023b). Na visão de Lütge (2021), a governança participativa não pode ser relegada a segundo plano, enquanto Taddeo (2023a) delinea propostas de modelos abrangentes que fundem a cooperação de múltiplos setores, de forma a assegurar que as inquietudes e aspirações da população sejam efetivamente captadas e incorporadas.

### **5.2.5. Componente 5: Sistema de Monitoramento e Adequação Regulatória (SMAR)**

Este componente garante a conformidade legal e a antecipação de futuras regulamentações.

- **Descrição:** Sistema de vigilância proativa que acompanha a evolução contínua da legislação e das normas de IA em níveis variados — nacional, regional e internacional — e promove a atualização permanente das práticas internas e dos sistemas de IA, garantindo a conformidade e atenuando os riscos jurídicos.
- **Fundamentação Teórica:** ajustar a inteligência artificial aos marcos regulatórios e às normas jurídicas é imprescindível para que ela exerça uma “conduta cidadã”. Captar as sutilezas da governança e das políticas regionais e nacionais revela-se vital,

conforme apontam as reflexões de Floridi (2023a, 2023b) sobre China, UE e Reino Unido; as análises de Wang (2024a, 2024b, 2022b, 2021a) referentes à China e à Lei de IA da UE; a perspectiva de Lütge (2022a) acerca da abordagem europeia; e as observações de Pagallo (2024a, 2017) acerca das barreiras legais. Na sua análise da Lei da IA, Wang (2024b) evidencia as repercussões da referida norma sobre as auditorias algorítmicas, apontando, assim, a urgência de as organizações se ajustarem às diretrizes estabelecidas.

### **5.2.6. Componente 6: Programa de Conscientização e Capacitação Ética em IA e Definição de Papéis (PCCEIA)**

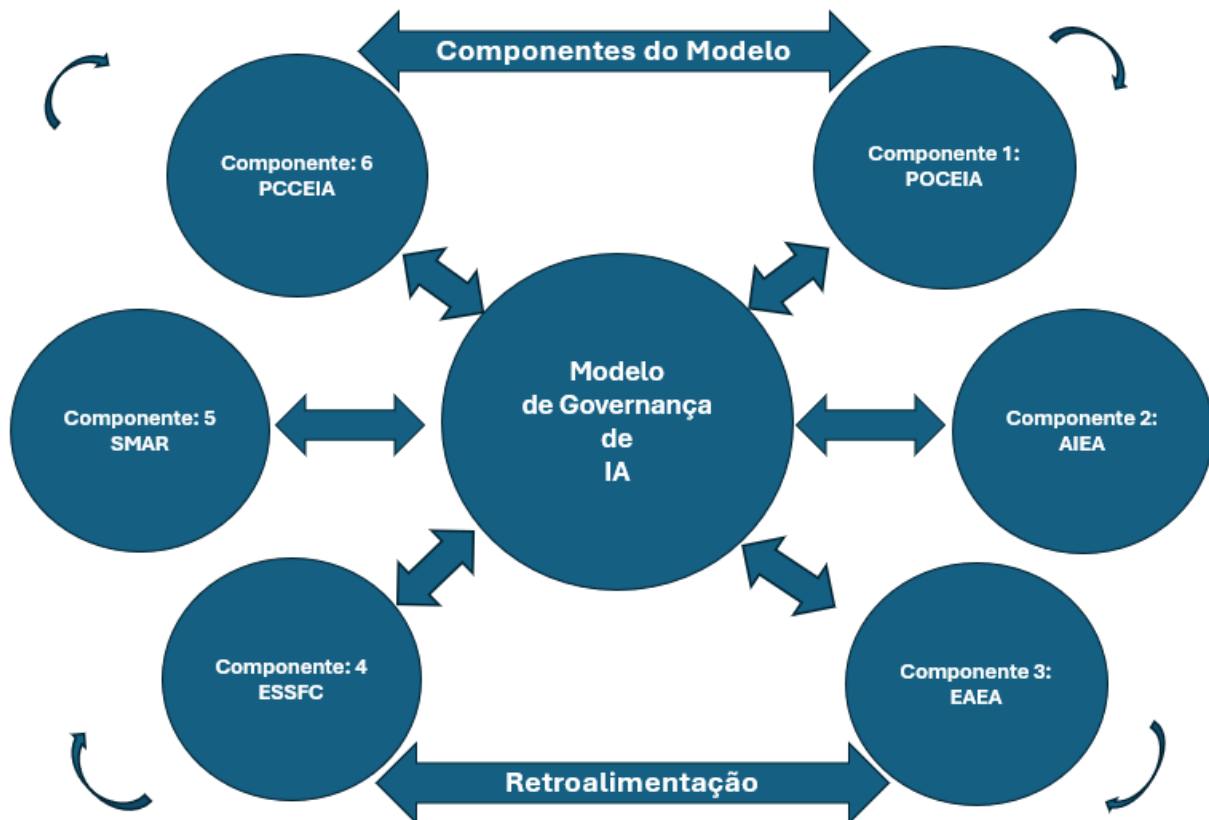
Este componente visa fomentar uma cultura organizacional ética.

- Descrição: Elaboração de programas de treinamento contínuo, workshops e diretrizes internas que visam educar e capacitar todos os colaboradores que lidam com IA, destacando os princípios éticos e as práticas de governança. Define papéis, responsabilidades e *accountabilities*, todos bem definidos, para assegurar a governança ética da IA na estrutura organizacional.
- Fundamentação Teórica: instaurar uma cultura organizacional ética funciona como construir um pilar vital. A necessidade de debater a inserção de princípios de inovação responsável e a construção de uma “cultura ética” (Floridi, 2022b, 2018a; Stahl, 2022b, 2024, 2021a, 2021b; Taddeo, 2025a) mostra-se imperativa. Floridi (2018a) cunhou o termo “soft ethics”, e Stahl (2021a, 2021b) posiciona-se como um ardente defensor da boa governança e da inovação responsável. Taddeo (2025a) defende que a governança incorpore o raciocínio ético, destacando a

importância dos funcionários e gestores ao deliberarem sobre decisões morais.

**A figura 4:** a seguir, tem-se o modelo de governança em IA, baseado em 6 componentes, conforme descrito anteriormente.

**Figura 4.** Modelo de governança em IA



**Fonte:** Elaborado pelos autores.

Este modelo de governança, concebido sob a ótica do Design Science Research e alicerçado numa densa base teórica e empírica da literatura sobre ética e governança da IA, funciona como um artefato prático para as organizações. Combinando princípios, processos e estruturas que introduzem valores desde a concepção, mitigam riscos específicos, garantem prestação de contas e transparência, envolvem os stakeholders, se adaptam ao panorama regulatório e fomentam uma cultura ética, o modelo capacita as organizações a irem além de uma visão meramente técnica da IA.

Ele disponibiliza a estrutura indispensável para que as organizações elaborem sistemas de IA que extrapolam o mero avanço tecnológico, enraizando-se, profundamente, em valores como justiça, equidade, transparência e respeito à dignidade humana, tudo isso guiado por uma verdadeira “conduta cidadã”. A adoção desse modelo simplificará a criação de IA que gere impactos positivos na sociedade, ao mesmo tempo em que reforça a confiança e a responsabilidade no ecossistema da IA.

## **6. CONCLUSÃO**

O trabalho investigou a necessidade cada vez mais evidente de incluir princípios éticos e estruturas de governança sólidas na criação e aplicação de sistemas de Inteligência Artificial (IA) tanto no setor público quanto no privado. A partir da intenção de mapear e examinar esses princípios, o estudo propôs um modelo prático, buscando responder à pergunta central: de que maneira as perspectivas de ética e governança podem apoiar, de forma efetiva, a construção de modelos de IA guiados por valores cidadãos.

A análise aprofundada da literatura apontou a complexidade e a natureza multifacetada das discussões sobre ética (com abordagens que vão de frameworks gerais a questões específicas, passando por auditorias éticas e debates narrativos) e sobre governança (abrangendo modelos genéricos, políticas regionais e nacionais, governança internacional, corporativa e auditorias como instrumento).

Este artefato se compõe de seis módulos interconectados: 1) Princípios Orientadores e Código de Ética de IA (POCEIA), que estabelecem a fundação moral; 2) Processo de Avaliação de Impacto

Ético da IA e Gerenciamento de Riscos (AIEA), focado na mitigação antecipada de riscos; 3) Estrutura de Auditoria Ética e Algorítmica (EAEA), que assegura responsabilidade e transparência; 4) Estrutura de Engajamento de Stakeholders e Feedback Contínuo (EESFC), garantindo a participação de diversas partes interessadas; 5) Sistema de Monitoramento e Adequação Regulatória (SMAR), que assegura a conformidade legal; e 6) Programa de Conscientização e Capacitação Ética em IA e Definição de Papéis (PCCEIA), que promove uma cultura organizacional ética.

O modelo proposto oferece um guia sistêmico e prático, capacitando as organizações a extrapolar a visão meramente técnica da IA. Ao integrar valores fundamentais desde a concepção, mitigar riscos específicos, estabelecer mecanismos claros de prestação de contas, promover o engajamento multissetorial, garantir a adaptação regulatória e fomentar uma cultura organizacional ética, o artefato desenvolvido assegura que os sistemas de IA sejam não apenas tecnologicamente avançados, mas também profundamente enraizados em princípios de justiça, equidade, transparência e respeito à dignidade humana.

Dessa forma, a IA se torna verdadeiramente orientada por condutas cidadãs, contribuindo positivamente para a sociedade e reforçando a confiança e a responsabilidade em todo o ecossistema de IA.

Para validação e aprimoramento contínuo do modelo, sugere-se que futuras pesquisas se concentrem na avaliação empírica do artefato em contextos organizacionais reais, conforme previsto na Fase 5 do processo de DSR. Tal etapa será crucial para verificar a utilidade e a eficácia do modelo em situações práticas e adaptá-lo às dinâmicas do mundo real.

## REFERÊNCIAS BIBLIOGRÁFICAS

Floridi, L. (2024a). Competing narratives in AI ethics: A defense of sociotechnical pragmatism. *AI & Society*. <https://doi.org/10.1007/s00146-024-02128-2>

Floridi, L. (2024b). Ten principles for responsible quantum innovation. *Quantum Science and Technology*. <https://doi.org/10.1088/2058-9565/ad3776>

Floridi, L. (2024c). The case for a broader approach to AI assurance: Addressing "hidden" harms in the development of artificial intelligence. *AI & Society*. <https://doi.org/10.1007/s00146-024-01950-y>

Floridi, L. (2024d). Smart cities: Reviewing the debate about their ethical implications. *AI & Society*. <https://doi.org/10.1007/s00146-022-01558-0>

Floridi, L. (2023a). Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *Information Society*. <https://doi.org/10.1080/01972243.2022.2124565>

Floridi, L. (2023b). Artificial intelligence regulation in the United Kingdom: A path to good governance and global leadership? *Internet Policy Review*. <https://doi.org/10.14763/2023.2.1709>

Floridi, L. (2023c). AI in support of the SDGs: Six recurring challenges and related opportunities identified through use cases. *Philosophical Studies Series*. [https://doi.org/10.1007/978-3-031-21147-8\\_2](https://doi.org/10.1007/978-3-031-21147-8_2)

Floridi, L. (2022a). The ethics of algorithms: Key problems and solutions. *AI & Society*. <https://doi.org/10.1007/s00146-021-01154-8>

Floridi, L. (2022b). Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry. *Frontiers in Computer Science*. <https://doi.org/10.3389/fcomp.2022.1068361>

Floridi, L. (2021a). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Philosophical Studies Series*. [https://doi.org/10.1007/978-3-030-81907-1\\_3](https://doi.org/10.1007/978-3-030-81907-1_3)

Floridi, L. (2021b). Ethics-based auditing to develop trustworthy AI. *Minds and Machines*. <https://doi.org/10.1007/s11023-021-09557-8>

Floridi, L. (2021c). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-021-00319-4>

Floridi, L. (2021d). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*. <https://doi.org/10.1007/s00146-020-00992-2>

Floridi, L. (2020a). The ethics of digital well-being: A thematic review. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-020-00175-8>

Floridi, L. (2020b). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-019-00165-5>

Floridi, L. (2018a). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*. <https://doi.org/10.1098/rsta.2018.0081>

Floridi, L., Cows, J., Beltrametti, M. A., Chatila, R., Chazerand, P., Dignum, V., Lütge, C., ... & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Guimarães, G. M. C.; Rocha Filho, G. P.; Marques, G. S.; Canedo, E. D. May I speak? Perceptions on ethical concerns and power while developing software in AI teams. In: SIMPÓSIO BRASILEIRO DE QUALIDADE DE SOFTWARE (SBQS), 24. , 2025, São José dos Campos/SP. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025 . p. 44-54. DOI: <https://doi.org/10.5753/sbqs.2025.13576>.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.

Lütge, C. (2023a). Investigating accountability in human-AI interaction: A philosophical and psychological perspective. *Frontiers in Psychology*, 14, 1073686. <https://doi.org/10.3389/fpsyg.2023.1073686>

Lütge, C. (2023b). Governing AI during a pandemic: Lessons from the COVID-19 crisis for responsible AI development. *Technology in Society*, 72, 102204. <https://doi.org/10.1016/j.techsoc.2023.102204>

Lütge, C. (2023c). From Trustworthy AI Principles to Practice: A Framework for Responsible AI Development and Deployment. *AI*,

4(4), 1046–1060. <https://doi.org/10.3390/ai4040046>

Lütge, C. (2022a). Governing AI: Attempting a European Approach. *Journal of European Public Policy*, 29(12), 2022–2040. <https://doi.org/10.1080/13501763.2022.2126515>

Lütge, C. (2021). AI4People: Ethical Guidelines for a Good AI Society. *International Journal of Technology and Human Interaction*, 17(1), Article 2. <https://doi.org/10.4018/IJT.20210101.oa2>

Mokander, J. (2024). Competing narratives in AI & society. *AI & Society*. <https://doi.org/10.1007/s00146-024-02128-2>

Mokander, J. (2021a). Ethics-Based Auditing to Minimize Algorithmic Risk. *Minds and Machines*, 31, 219–234. <https://doi.org/10.1007/s11023-021-09557-8>

Mokander, J. (2021b). Ethics-Based Auditing of Science and Engineering. *Science and Engineering Ethics*, 27(3), 22–34. <https://doi.org/10.1007/s11948-021-00319-4>

Monteiro, W. R. & Ayrosa, E. A. T. (2025). The AI-Based Leadership Method: fostering symbiotic integration of Artificial Intelligence into management practices. *Revista Gestão & Tecnologia*, vol 25, nº3, p: 263-285. <https://doi.org/10.20397/2177-6652/2025.v25i3.2810>

Pagallo, U. (2024a). The underuse of AI in the health and technology sector: ethical and legal barriers. *AI and Ethics*, 4(2), 297–308. <https://doi.org/10.1007/s12553-023-00806-7>

Pagallo, U. (2024b). Opening the ‘black box’ of human-AI communication: a regulatory agenda for AI transparency.

Communication Research and Practice.

<https://doi.org/10.1080/22041451.2024.2346415>

Pagallo, U. (2024c). The new laws of outer space: ethics, legislation, and public policy in orbit. *Artificial Intelligence and Law*.

<https://doi.org/10.1007/s10506-024-09377-0>

Pagallo, U. (2021). An Ethical Framework for AI. In *Philosophical Studies* (pp. 45–61). [https://doi.org/10.1007/978-3-030-81907-1\\_3](https://doi.org/10.1007/978-3-030-81907-1_3)

Pagallo, U. (2017). AI Features, Legal Futures. *Sistemi Intelligenti*, 1, 13–29. <https://doi.org/10.1422/88512>

Roberts, H. (2024). Global AI Governance: Bridging international affairs and AI ethics. *International Affairs*, 100(4), 1641–1658.

<https://doi.org/10.1093/ia/iiae073>

Roberts, H., & Cowls, J. (2023). Governing Artificial Intelligence: Interdisciplinary Perspectives on Information Society. *The Information Society*, 39(1), 1–4.

<https://doi.org/10.1080/01972243.2022.2124565>

Sordi, J. O. De; Meireles, M.; Sanches, C. Design science aplicada às pesquisas em administração: reflexões a partir do recente histórico de publicações internacionais. (2011). *INMR - Innovation & Management Review*,

8(1), 10–36. <https://revistas.usp.br/rai/article/view/79201>

Stahl, B. C. (2024). Towards trustworthy AI & society. *AI & Society*.

<https://doi.org/10.1007/s00146-024-02082-z>

Stahl, B. (2022a). Responsible Innovation. *International Journal of Information Management*, 62, 102441. <https://doi.org/10.1016/j.ijinfomgt.2021.102441>

Stahl, B. (2022b). A European Agency for Artificial Intelligence. *Computer Law & Security Review*, 46, 105661. <https://doi.org/10.1016/j.clsr.2022.105661>

Stahl, B. (2021a). Good Governance as a Response to AI Risks. *Interdisciplinary Science Reviews*, 46(1), 105–124. <https://doi.org/10.1080/03080188.2020.1840220>

Stahl, B. (2021b). Artificial Intelligence for Good Governance. *Journal of Business Research*, 124, 1–12. <https://doi.org/10.1016/j.jbusres.2020.11.030>

Taddeo, M. (2025a). Ethical governance of AI: the need for an ethics-based approach. *AI & Society*. Advance online publication. <https://doi.org/10.1007/s00146-024-01866-7>

Taddeo, M. (2024a). Global AI governance: Balancing innovation, ethics and security. *International Affairs*, 100(4), 1639-1658. <https://doi.org/10.1093/ia/iiae073>

Taddeo, M. (2024b). Open Source Intelligence (OSINT): an ethical taxonomy. *AI & Society*, 39(3), 1187-1196. <https://doi.org/10.1007/s00146-023-01628-x>

Taddeo, M. (2024c). The case for a broader approach to AI ethics: A reply to Floridi. *AI & Society*. Advance online publication. <https://doi.org/10.1007/s00146-024-01950-y>

Taddeo, M. (2024d). Introduction. In M. Taddeo (Ed.), *The ethics of artificial intelligence*. Oxford University Press.  
<https://doi.org/10.1093/oso/9780197745441.001.0>

Taddeo, M. (2023a). Governing artificial intelligence: A comprehensive ethical framework. *The Information Society*, 39(3), 200-212. <https://doi.org/10.1080/01972243.2022.2124565>

Taddeo, M. (2023b). AI in Support of the SDGs: Ethical Considerations. In E. A. et al. (Eds.), *Philosophical Studies in Artificial Intelligence and Its Social and Ethical Implications* (pp. 15–30). Springer, Cham.  
[https://doi.org/10.1007/978-3-031-21147-8\\_2](https://doi.org/10.1007/978-3-031-21147-8_2)

Taddeo, M. (2023c). Artificial intelligence and the future of human agency. *Internet Policy Review*, 12(2).  
<https://doi.org/10.14763/2023.2.1709>

Taddeo, M. (2022a). Apropos data sharing: A response to Knoppers and Deschenes. *DNA and Cell Biology*, 41(4), 316-318.  
<https://doi.org/10.1089/dna.2021.0501>

Taddeo, M. (2022b). The ethics of algorithms: a taxonomy for the responsible design and implementation of AI. *AI & Society*, 37(1), 1–11.  
<https://doi.org/10.1007/s00146-021-01154-8>

Taddeo, M., & Floridi, L. (2023). The AI gambit: leveraging artificial intelligence for trustworthy digital ecosystems. *AI & Society*, 38(1), 1–11.  
<https://doi.org/10.1007/s00146-021-01294-x>

Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of algorithms: a new governance mechanism for trustworthy AI. *Science and*

Engineering Ethics, 27(5), 52. <https://doi.org/10.1007/s11948-021-00319-4>

Taddeo, M., & Floridi, L. (2020). The ethics of digital well-being. *Science and Engineering Ethics*, 26(3), 1279-1300. <https://doi.org/10.1007/s11948-020-00175-8>

Wang, M. Y., Huang, H. H., Wu, Y. S., Chu, C. L., Lin, R. J., Lu, H. F., ... & Su, S. B. (2023). Guidelines for Ethical Use of Artificial Intelligence in Clinical Practice: A Consensus Statement. *Journal of Medical Internet Research*, 25, e46089. <https://doi.org/10.2196/46089>

Wang, Y. (2024a). Ethics and Governance of Artificial Intelligence in Medicine in China: Progress and Challenges. *Chinese Medical Ethics*, 41(9), 1109-1113. <https://doi.org/10.12026/j.issn.1001-8565.2024.09>

Wang, Y. (2024b). "Game Changer": The AI Act and Algorithmic Audits in Financial Services. *AI & Society*. <https://doi.org/10.1007/s00146-024-02027-6>

Wang, Y. (2023a). Artificial Intelligence and Sustainability: An Overview of Opportunities and Challenges. *Sustainability*, 15(6), 5174. <https://doi.org/10.3390/su15064796>

Wang, Y. (2022a). Shifting artificial intelligence (AI) ethics from principle to practice: An ethical framework for social science research on AI. *Social Science & Medicine*, 315, 114782. <https://doi.org/10.1016/j.socscimed.2022.114782>

Wang, Y. (2022b). Ethics and Governance of AI in China: A Review of Policy and Practice. *Bulletin of the Chinese Academy of Sciences*, 37(5), 527-535. <https://doi.org/10.16418/j.issn.1000-3045.2022050>

Wang, Y. (2021a). Ethics Guidelines for Artificial Intelligence in China: An Analysis of the WTO Regulatory Landscape. *China WTO Review*, 7(2), 203-222. <https://doi.org/10.14330/cwr.2021.7.2.03>

---

<sup>1</sup> Universidade de Brasília (UnB), Departamento de Ciência da Computação. ORCID: <https://orcid.org/0000-0001-5900-3749>. E-mail: [acesse o artigo original para visualizar o e-mail](#)

<sup>2</sup> Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). ORCID: <https://orcid.org/0000-0003-2833-9367>. E-mail: [acesse o artigo original para visualizar o e-mail](#)

<sup>3</sup> Universidade de Brasília (UnB), Departamento de Engenharia Elétrica, ORCID: <https://orcid.org/0000-0001-7100-7304>. E-mail: [acesse o artigo original para visualizar o e-mail](#)

<sup>4</sup> Universidade de Brasília (UnB), Departamento de Ciência da Computação, ORCID: <https://orcid.org/0000-0002-2159-339X>. E-mail: [acesse o artigo original para visualizar o e-mail](#)