

Governança Ética da Inteligência Artificial: uma análise comparativa entre diretrizes éticas e marcos regulatórios aplicáveis a sistemas de IA de alto risco

Ethical Governance of Artificial Intelligence: a comparative analysis of ethical guidelines and regulatory frameworks applicable to high-risk AI systems

ÁREA TEMÁTICA: Ética e Regulação da Inteligência Artificial

Leandro Nascimento de Oliveira, Universidade de Brasília, Brasil, leandroo65@gmail.com

Edna Dias Canedo, Universidade de Brasília, Brasil, ednacanedo@unb.br

RESUMO

A crescente adoção da Inteligência Artificial (IA) em setores como saúde, segurança pública, finanças e defesa tem impulsionado debates sobre os limites éticos e jurídicos de sua utilização. Embora organismos internacionais e governos tenham desenvolvido diretrizes e instrumentos regulatórios voltados ao uso responsável da IA, ainda são observadas divergências e lacunas entre princípios éticos e exigências normativas, sobretudo em aplicações consideradas de alto risco. Situações envolvendo vigilância em massa, sistemas de armas autônomas e processos decisórios automatizados capazes de afetar direitos fundamentais evidenciam a necessidade de uma governança que concilie inovação, proteção de direitos e responsabilidade institucional. Neste contexto, o presente estudo analisa as convergências e lacunas existentes entre diretrizes éticas e marcos regulatórios nacionais e internacionais relacionados à governança da Inteligência Artificial. Para tanto, adota-se uma abordagem qualitativa, de caráter exploratório e descritivo, baseada em revisão bibliográfica e análise documental de normas e recomendações, incluindo o EU AI Act, a Lei Geral de Proteção de Dados (LGPD), o General Data Protection Regulation (GDPR), os Princípios de Inteligência Artificial da OCDE e a Recomendação da UNESCO sobre a Ética da Inteligência Artificial. A análise busca compreender como esses instrumentos tratam aplicações de alto risco, bem como os impactos éticos, jurídicos e reputacionais decorrentes da ausência de práticas adequadas de governança. Ao final, são discutidos os desafios para a harmonização entre ética e regulação e sua importância para o desenvolvimento de sistemas de IA alinhados à proteção dos direitos fundamentais.

Palavras-chave: Inteligência Artificial; Governança de IA; Ética em IA; Regulação da IA; Direitos Fundamentais; Sistemas de Alto Risco.

ABSTRACT

The growing adoption of Artificial Intelligence (AI) in sectors such as healthcare, public security, finance, and defense has spurred debates regarding the ethical and legal boundaries of its use. Although international bodies and governments have developed guidelines and regulatory instruments aimed at the responsible use of AI, discrepancies and gaps persist between ethical principles and regulatory requirements, particularly in applications deemed high-risk. Situations involving mass surveillance, autonomous weapon systems, and automated decision-making processes capable of affecting fundamental rights highlight the need for a governance framework that balances innovation, rights protection, and institutional accountability. In this context, this study analyzes the convergences and gaps between ethical guidelines and national and international regulatory frameworks regarding AI governance. To this end, a qualitative, exploratory, and descriptive approach is adopted, based on a literature review and document analysis of standards and recommendations—including the EU AI Act, the General Data Protection Law (LGPD), the General Data Protection Regulation (GDPR), the OECD AI Principles, and the UNESCO Recommendation on the Ethics of Artificial Intelligence. The analysis seeks to understand how these instruments address high-risk applications, as well as the ethical, legal, and reputational impacts resulting from the absence of adequate governance practices. Finally, the study discusses the challenges of harmonizing ethics and regulation and the importance of this alignment for developing AI systems that uphold fundamental rights.

1. INTRODUÇÃO

A Inteligência Artificial (IA) tem assumido papel cada vez mais relevante nas transformações tecnológicas contemporâneas, ampliando sua presença em diferentes setores da sociedade e influenciando processos de tomada de decisão anteriormente desempenhados exclusivamente por seres humanos. Aplicações baseadas em IA são empregadas em áreas como saúde, segurança pública, sistema financeiro, educação, mobilidade e defesa, contribuindo para ganhos de eficiência operacional, capacidade analítica e automação de processos. Nesse contexto, a IA é frequentemente apontada como uma das tecnologias com maior potencial para impulsionar inovação, desenvolvimento econômico e bem-estar social (Russell; Norvig, 2021).

Entretanto, a crescente adoção dessas tecnologias também tem evidenciado desafios éticos, jurídicos e sociais relacionados ao seu desenvolvimento e utilização. Questões envolvendo privacidade, discriminação algorítmica, falta de transparência nos processos decisórios automatizados, segurança da informação e responsabilização por danos decorrentes do uso de sistemas inteligentes têm ocupado posição central nos debates acadêmicos, regulatórios e institucionais (Heilinger, 2022). Tais preocupações tornam-se ainda mais relevantes em contextos considerados de alto risco, nos quais decisões automatizadas podem impactar diretamente direitos fundamentais, como ocorre em sistemas de vigilância, reconhecimento facial, concessão de crédito, recrutamento de pessoal e aplicações voltadas à segurança pública.

A literatura recente destaca que a transparência constitui um dos principais pilares para a construção de sistemas de IA confiáveis e socialmente legítimos. Contudo, muitos modelos computacionais operam por meio de mecanismos complexos de aprendizado de máquina que dificultam a compreensão sobre como determinadas decisões são produzidas, fenômeno frequentemente denominado de “caixa-preta” (black box). Essa limitação compromete a capacidade de auditoria, supervisão e responsabilização dos agentes envolvidos, além de dificultar a identificação de vieses e potenciais violações de direitos (Krook et al., 2025). Segundo os autores, a crescente preocupação com a opacidade dos sistemas inteligentes tem impulsionado iniciativas regulatórias em diferentes jurisdições, especialmente na União Europeia e no Reino Unido, que passaram a incorporar requisitos de transparência, explicabilidade e governança em seus modelos regulatórios.

Em resposta a esses desafios, organismos internacionais, governos e instituições privadas vêm desenvolvendo princípios éticos, diretrizes de governança e instrumentos normativos voltados à promoção do uso responsável da Inteligência Artificial. Entre os principais referenciais destacam-se os Princípios de Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a Recomendação da UNESCO sobre a Ética da Inteligência Artificial, o Regulamento Europeu de Inteligência Artificial (EU AI Act), o Regulamento Geral sobre a Proteção de Dados (GDPR) e, no contexto brasileiro, a Lei Geral de Proteção de Dados Pessoais (LGPD). Esses instrumentos compartilham preocupações relacionadas à transparência, à supervisão humana, à mitigação de riscos e à proteção dos direitos fundamentais, embora apresentem diferenças significativas quanto ao alcance, à aplicabilidade e aos mecanismos de fiscalização.

Apesar dos avanços observados no campo regulatório, ainda existem lacunas entre os princípios éticos amplamente defendidos pela literatura e as exigências efetivamente previstas nos instrumentos normativos. A própria literatura especializada aponta que a adoção de mecanismos regulatórios, por si só, não é suficiente para assegurar práticas responsáveis de desenvolvimento e utilização da IA, sendo necessária a incorporação de valores éticos e estruturas de governança capazes de orientar decisões organizacionais ao longo de todo o ciclo de vida dos sistemas inteligentes (Pahune et al., 2025). Nesse sentido, a ausência de mecanismos adequados de governança pode ampliar riscos relacionados a vieses algorítmicos, uso indevido de dados, falhas de segurança, violações de privacidade e descumprimento de obrigações legais. Diante desse cenário, esta pesquisa busca responder às seguintes questões: de que forma diretrizes éticas e marcos regulatórios abordam a governança responsável da Inteligência Artificial? Quais são as principais convergências e divergências entre esses instrumentos? E quais impactos podem

decorrer da ausência de princípios éticos na formulação de políticas organizacionais voltadas ao desenvolvimento e à utilização de sistemas de IA?

O objetivo geral deste estudo consiste em analisar as convergências e lacunas existentes entre diretrizes éticas e marcos regulatórios relacionados à governança da Inteligência Artificial, com ênfase em aplicações consideradas de alto risco. Como objetivos específicos, pretende-se: identificar os princípios presentes nas principais diretrizes éticas e legislações sobre IA; comparar os mecanismos de governança previstos nesses instrumentos; analisar os riscos éticos, jurídicos e organizacionais decorrentes da ausência de práticas adequadas de governança; e discutir possíveis contribuições para o fortalecimento de modelos de IA alinhados à proteção dos direitos fundamentais.

A relevância da pesquisa decorre da crescente dependência de sistemas inteligentes em processos decisórios com elevado impacto social, econômico e institucional. Em um contexto marcado pela rápida evolução tecnológica e pela consolidação de novos modelos regulatórios em diferentes países, compreender as relações entre ética, governança e regulação torna-se fundamental para o desenvolvimento de sistemas de IA que conciliem inovação, segurança jurídica, transparência e respeito aos direitos fundamentais. Assim, a presente pesquisa pretende contribuir para o debate acadêmico e profissional sobre governança da Inteligência Artificial, oferecendo uma análise crítica das convergências e lacunas existentes entre os referenciais éticos e regulatórios atualmente em vigor.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Ética em Inteligência Artificial

O avanço da Inteligência Artificial (IA) trouxe consigo uma série de desafios que extrapolam os aspectos tecnológicos e alcançam dimensões éticas, jurídicas e sociais. À medida que sistemas inteligentes passaram a desempenhar funções relacionadas à tomada de decisão em áreas como saúde, segurança pública, mercado financeiro e administração pública, aumentou também a preocupação com os impactos dessas tecnologias sobre direitos fundamentais, privacidade, igualdade e autonomia dos indivíduos. Nesse contexto, a discussão sobre ética em Inteligência Artificial consolidou-se como um dos principais eixos do debate internacional acerca do desenvolvimento e da utilização responsável dessas tecnologias.

De maneira geral, a ética em Inteligência Artificial (*Artificial Intelligence Ethics* ou *AI Ethics*) pode ser compreendida como o conjunto de princípios e valores que orientam o desenvolvimento, a implementação e o uso de sistemas inteligentes, buscando assegurar que essas tecnologias estejam alinhadas aos direitos humanos, aos valores democráticos e ao interesse público. Para Floridi e Cowls (2019), a ética em IA constitui um campo interdisciplinar voltado à construção de parâmetros capazes de orientar decisões tecnológicas socialmente responsáveis, conciliando inovação, proteção dos direitos fundamentais e bem-estar coletivo.

Embora diferentes organismos internacionais adotem abordagens próprias, observa-se um consenso em torno de alguns princípios considerados essenciais para uma IA confiável (*trustworthy AI*). Entre eles destacam-se a transparência, a responsabilização (*accountability*), a justiça (*fairness*), a explicabilidade (*explainability*), a supervisão humana, a segurança, a privacidade e a não discriminação. Esses princípios aparecem de forma recorrente em documentos internacionais, como os Princípios sobre Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a Recomendação da UNESCO sobre a Ética da Inteligência Artificial e o Regulamento Europeu de Inteligência Artificial (EU AI Act), ainda que com diferentes níveis de detalhamento e aplicação.

Figura 1: Convergência dos princípios éticos da Inteligência Artificial

CONVERGÊNCIA DOS PRINCÍPIOS ÉTICOS DA INTELIGÊNCIA ARTIFICIAL

Síntese comparativa dos principais referenciais internacionais



REFERENCIAIS INTERNACIONAIS CONSIDERADOS					
OECD AI Principles (2019)	HLEG – Ethics Guidelines for Trustworthy AI (2019)	UNESCO Recommendation on the Ethics of AI (2021)	NIST NIST AI RMF 1.0 (2023)	ISO/IEC 42001:2023	EU AI Act (2024)
✓ Transparência e explicabilidade ✓ Accountability ✓ Justiça e não discriminação ✓ Privacidade (tratamento de dados) ✓ Robustez, segurança e proteção ✓ Crescimento inclusivo e bem-estar	✓ Supervisão humana ✓ Robustez técnica e segurança ✓ Privacidade e governança de dados ✓ Transparência ✓ Diversidade, não discriminação e equidade ✓ Bem-estar social e ambiental ✓ Accountability	✓ Direitos humanos e dignidade ✓ Inclusão e diversidade ✓ Igualdade e não discriminação ✓ Privacidade e proteção de dados ✓ Transparência e explicabilidade ✓ Segurança e proteção ✓ Responsabilidade e prestação de contas ✓ Sustentabilidade	✓ Mapeamento e gestão de riscos ✓ Transparência ✓ Explicabilidade ✓ Privacidade ✓ Segurança e resiliência ✓ Mitigação de vieses ✓ Accountability	✓ Gestão da IA responsável ✓ Governança e liderança ✓ Transparência ✓ Gestão de riscos ✓ Privacidade e proteção de dados ✓ Accountability ✓ Melhoria contínua	✓ Transparência ✓ Supervisão humana ✓ Gestão de riscos ✓ Governança de dados ✓ Robustez e precisão ✓ Segurança ✓ Accountability

Nota: Os princípios apresentados representam a convergência dos elementos mais recorrentes entre os referenciais analisados. A redação foi adaptada pelo autor para fins de organização e síntese.
Fonte: Elaborado pelo autor (2026), com base em OECD (2019); HLEG (2019); UNESCO (2021); NIST AI RMF 1.0 (2023); ISO/IEC 42001:2023; EU AI Act (2024).

A **transparência** é um dos princípios mais discutidos na literatura especializada. Sua finalidade é possibilitar que os critérios utilizados pelos sistemas de IA para produzir recomendações, classificações ou decisões sejam compreendidos por usuários, desenvolvedores, auditores e órgãos reguladores. Em modelos baseados em aprendizado de máquina, especialmente aqueles que utilizam técnicas de aprendizado profundo, o funcionamento interno frequentemente apresenta elevado grau de complexidade, dificultando a compreensão sobre como determinadas conclusões foram alcançadas. Esse fenômeno, conhecido como *black box*, representa um dos principais desafios para a adoção responsável da IA, pois limita a capacidade de auditoria, reduz a confiança dos usuários e dificulta a identificação de possíveis vieses ou erros de funcionamento.

Outro princípio amplamente reconhecido é a **accountability**, normalmente traduzida como responsabilização ou prestação de contas. No contexto da Inteligência Artificial, esse princípio estabelece que os agentes envolvidos no desenvolvimento, implantação e utilização dos sistemas devem assumir responsabilidade pelos impactos decorrentes de seu funcionamento. Isso implica a existência de mecanismos que permitam identificar responsáveis, registrar decisões, possibilitar auditorias e garantir que eventuais danos possam ser investigados e corrigidos. A responsabilização constitui elemento essencial para promover confiança e segurança jurídica na utilização dessas tecnologias.

A **fairness**, frequentemente traduzida como equidade ou justiça algorítmica, busca assegurar que sistemas de IA não produzam tratamentos discriminatórios em razão de características como raça, gênero, idade, condição socioeconômica ou qualquer outro fator protegido por lei. Conforme observa O'Neil (2016), algoritmos treinados com bases de dados enviesadas podem reproduzir ou até ampliar desigualdades existentes na sociedade, conferindo aparência de neutralidade a decisões essencialmente discriminatórias. Dessa forma, promover equidade exige não apenas qualidade dos dados utilizados, mas também avaliação contínua dos resultados produzidos pelos modelos ao longo de sua utilização.

A **explicabilidade** (*explainability*) também ocupa posição de destaque nas discussões sobre

ética em IA. Embora frequentemente associada à transparência, trata-se de um conceito distinto. Enquanto a transparência está relacionada à disponibilização de informações sobre o funcionamento do sistema, a explicabilidade refere-se à capacidade de apresentar justificativas compreensíveis para as decisões produzidas pela Inteligência Artificial. Esse princípio torna-se particularmente relevante em aplicações de alto impacto, nas quais decisões automatizadas podem influenciar o acesso a crédito, oportunidades de emprego, tratamentos médicos ou serviços públicos. Conforme destacam Krook et al. (2025), a ausência de mecanismos de explicação reduz a confiança nos sistemas inteligentes e dificulta sua avaliação por usuários e autoridades competentes.

Embora esses princípios sejam amplamente reconhecidos pela comunidade científica e por organismos internacionais, sua efetiva implementação ainda representa um desafio para organizações públicas e privadas. Em muitos casos, documentos de caráter ético estabelecem valores e recomendações gerais, mas não definem mecanismos concretos para assegurar sua aplicação na prática. Essa constatação evidencia que a adoção de princípios éticos, por si só, não é suficiente para garantir o desenvolvimento e a utilização responsável da Inteligência Artificial, tornando necessária a existência de estruturas organizacionais capazes de transformar esses princípios em práticas efetivas. Essa discussão será aprofundada na próxima seção, dedicada ao estudo da Governança da Inteligência Artificial.

2.2 Governança da Inteligência Artificial

O crescimento da utilização da Inteligência Artificial em atividades estratégicas ampliou a necessidade de mecanismos capazes de assegurar que essas tecnologias sejam desenvolvidas, implementadas e utilizadas de maneira responsável. Nesse contexto, a governança da Inteligência Artificial (*Artificial Intelligence Governance* ou *AI Governance*) consolidou-se como um campo de estudo voltado à criação de estruturas organizacionais, instrumentos regulatórios e processos de gestão destinados a reduzir riscos, promover a conformidade normativa e garantir que sistemas inteligentes permaneçam alinhados aos objetivos institucionais e aos direitos fundamentais.

Embora o conceito ainda esteja em evolução, observa-se relativo consenso na literatura quanto à sua abrangência. Dafoe (2018) define a governança da IA como o conjunto de instituições, políticas, processos e mecanismos voltados a orientar o desenvolvimento e a utilização da Inteligência Artificial em benefício da sociedade, considerando aspectos políticos, econômicos, jurídicos, militares e éticos. Sob essa perspectiva, a governança não se limita ao cumprimento de normas legais, mas envolve a construção de capacidades institucionais que permitam administrar os riscos decorrentes da rápida evolução tecnológica.

Definição semelhante é apresentada por Mäntymäki et al. (2022), citados por Kim et al. (2025), para quem a governança da IA corresponde a um sistema integrado de regras, práticas, processos e instrumentos tecnológicos destinados a assegurar que o uso da Inteligência Artificial permaneça alinhado aos objetivos estratégicos da organização, às exigências legais e aos princípios éticos adotados institucionalmente. Essa abordagem amplia a compreensão da governança ao incorporá-la como parte da gestão organizacional, integrando aspectos técnicos, jurídicos e administrativos durante todo o ciclo de vida dos sistemas de IA.

Sob a perspectiva organizacional, a governança da IA pode ser compreendida como uma extensão das práticas tradicionais de governança corporativa. Assim como ocorre na governança de tecnologia da informação e na governança de dados, a adoção de sistemas inteligentes exige definição clara de papéis e responsabilidades, estabelecimento de políticas internas, mecanismos de supervisão, auditoria, monitoramento contínuo e processos destinados à mitigação de riscos. O objetivo é assegurar que decisões automatizadas permaneçam compatíveis com a estratégia institucional, respeitem os requisitos legais aplicáveis e preservem os direitos dos indivíduos potencialmente afetados.

Entre os componentes essenciais da governança destacam-se as estruturas organizacionais responsáveis pela supervisão do uso da IA. A literatura especializada recomenda a criação de comitês

multidisciplinares ou instâncias equivalentes, compostos por profissionais das áreas de tecnologia, segurança da informação, jurídico, proteção de dados, gestão de riscos, compliance e negócio. Esses grupos desempenham papel fundamental na definição de políticas internas, avaliação de riscos, aprovação de projetos, acompanhamento do desempenho dos sistemas inteligentes e tratamento de incidentes relacionados ao uso da IA.

Outro elemento indispensável refere-se aos programas de **compliance**. Embora tradicionalmente associados ao cumprimento de requisitos legais e regulatórios, programas de compliance voltados à Inteligência Artificial possuem alcance mais amplo. Além de verificar conformidade com legislações como a Lei Geral de Proteção de Dados (LGPD) e o Regulamento Europeu de Inteligência Artificial (EU AI Act), esses programas estabelecem procedimentos para documentação dos modelos, controle de qualidade dos dados utilizados, registro das decisões automatizadas, auditorias periódicas, capacitação de equipes e mecanismos destinados à prevenção de riscos éticos e operacionais.

Nesse contexto, a gestão de riscos ocupa posição central na governança da IA. Diferentemente de abordagens regulatórias estritamente prescritivas, estudos recentes defendem modelos baseados na identificação, avaliação, tratamento e monitoramento contínuo dos riscos associados à Inteligência Artificial. Coglianese e Crum (2025) argumentam que tecnologias caracterizadas por elevado dinamismo, como a IA, dificilmente podem ser reguladas por meio de regras rígidas e previamente definidas. Em substituição ao modelo tradicional de *guardrails* (barreiras regulatórias fixas), os autores propõem uma abordagem baseada em *management-based regulation*, fundamentada na supervisão humana permanente, na melhoria contínua e na responsabilidade institucional ao longo de todo o ciclo de vida dos sistemas inteligentes.

Essa perspectiva aproxima a governança da IA dos modelos contemporâneos de gestão de riscos organizacionais. Em vez de buscar eliminar completamente os riscos decorrentes da utilização da Inteligência Artificial — objetivo praticamente inalcançável diante da velocidade de evolução tecnológica —, procura-se estabelecer processos capazes de identificar vulnerabilidades, avaliar seus impactos e implementar controles proporcionais ao nível de risco apresentado por cada aplicação.

Essa lógica encontra respaldo no próprio EU AI Act, considerado atualmente o principal marco regulatório sobre Inteligência Artificial. O regulamento europeu adota uma abordagem baseada em risco (*risk-based approach*), classificando os sistemas de IA conforme seu potencial de impacto sobre direitos fundamentais, segurança e saúde. Sistemas considerados de risco inaceitável são proibidos, enquanto aplicações classificadas como de alto risco permanecem autorizadas, desde que atendam a requisitos específicos relacionados à gestão de riscos, documentação técnica, qualidade dos dados, transparência, supervisão humana e monitoramento contínuo.

Nos últimos anos, diferentes modelos de **AI Governance Frameworks** foram desenvolvidos com o objetivo de auxiliar organizações na implementação dessas práticas. Entre os referenciais mais relevantes destacam-se os Princípios sobre Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a Recomendação da UNESCO sobre a Ética da Inteligência Artificial, o NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0), publicado pelo National Institute of Standards and Technology, e a norma internacional ISO/IEC 42001:2023, primeiro padrão internacional voltado especificamente para Sistemas de Gestão de Inteligência Artificial.

Embora possuam enfoques distintos, esses referenciais apresentam elementos convergentes. Em geral, todos recomendam que organizações estabeleçam estruturas formais de governança, processos permanentes de gestão de riscos, mecanismos de supervisão humana, políticas de transparência, auditoria, prestação de contas e melhoria contínua. Também enfatizam que a governança da IA não deve ser compreendida como responsabilidade exclusiva das equipes técnicas, mas como compromisso institucional compartilhado entre diferentes áreas da organização.

Observa-se, portanto, que a governança da Inteligência Artificial representa uma evolução das

discussões exclusivamente éticas ou regulatórias. Seu propósito consiste em transformar princípios abstratos em práticas organizacionais capazes de orientar decisões concretas, reduzir riscos e fortalecer a confiança na utilização de sistemas inteligentes. Mais do que assegurar conformidade legal, modelos eficazes de governança contribuem para que a inovação tecnológica ocorra de forma responsável, transparente e compatível com a proteção dos direitos fundamentais, aspecto que se mostra particularmente relevante em aplicações classificadas como de alto risco.

2.3 Contextos sensíveis e aplicações de alto risco

Embora a Inteligência Artificial tenha ampliado significativamente sua capacidade de apoiar processos decisórios e otimizar atividades em diversos setores, determinadas aplicações apresentam potencial elevado para afetar direitos fundamentais, segurança pública e estabilidade institucional. Em razão dessas características, a literatura especializada e os marcos regulatórios mais recentes passaram a classificá-las como aplicações de alto risco, exigindo mecanismos mais rigorosos de supervisão, transparência, gestão de riscos e responsabilização.

A adoção dessa perspectiva tornou-se especialmente evidente com a publicação do Regulamento Europeu de Inteligência Artificial (EU AI Act), que introduziu um modelo regulatório baseado no risco (*risk-based approach*). Nesse modelo, o nível de exigência imposto aos desenvolvedores e usuários varia de acordo com os possíveis impactos produzidos pelos sistemas sobre a saúde, a segurança e os direitos fundamentais das pessoas. Sistemas considerados de risco inaceitável são proibidos, enquanto aplicações classificadas como de alto risco permanecem permitidas desde que atendam a requisitos específicos de governança, documentação técnica, supervisão humana, monitoramento contínuo e gestão de riscos.

Essa classificação reflete uma mudança importante na forma de compreender os riscos associados à Inteligência Artificial. Em vez de concentrar a atenção exclusivamente na tecnologia empregada, os modelos regulatórios contemporâneos passaram a avaliar os impactos sociais decorrentes de sua utilização. Assim, quanto maior o potencial de interferência sobre direitos individuais e coletivos, maior tende a ser a necessidade de mecanismos de controle, auditoria e prestação de contas.

2.3.1 Vigilância em massa

Entre as aplicações que despertam maior preocupação encontram-se os sistemas de vigilância em massa baseados em Inteligência Artificial. A integração entre técnicas de reconhecimento facial, análise comportamental, processamento de grandes volumes de dados e monitoramento em tempo real ampliou significativamente a capacidade de Estados e organizações privadas acompanharem atividades individuais em espaços físicos e digitais.

Embora essas tecnologias possam contribuir para investigações criminais, prevenção de fraudes e proteção da segurança pública, seu emprego indiscriminado levanta importantes questionamentos relacionados à privacidade, à liberdade de expressão, ao direito de reunião e à proteção de dados pessoais. A utilização permanente de mecanismos de vigilância pode produzir efeitos inibidores sobre comportamentos sociais, reduzindo a autonomia dos indivíduos e ampliando possibilidades de controle estatal ou corporativo.

Dafoe (2018) observa que sistemas avançados de IA podem fortalecer modelos de vigilância capazes de concentrar poder político e institucional em níveis sem precedentes, tornando mais complexa a preservação de direitos fundamentais em sociedades democráticas. Entre os riscos apontados pelo autor destacam-se o monitoramento permanente da população, a manipulação comportamental e a utilização de sistemas inteligentes para ampliar mecanismos de repressão e controle social.

Essas preocupações explicam a posição adotada pelo EU AI Act ao restringir severamente determinadas modalidades de identificação biométrica remota em tempo real realizadas em espaços públicos, especialmente quando destinadas à vigilância indiscriminada da população. O regulamento

reconhece que aplicações dessa natureza podem representar riscos incompatíveis com a proteção da dignidade humana, da privacidade e de outros direitos fundamentais.

2.3.2 Decisões automatizadas

Outro contexto considerado particularmente sensível refere-se à utilização de sistemas de Inteligência Artificial para apoiar ou substituir decisões humanas em processos que produzem efeitos relevantes sobre a vida das pessoas. Atualmente, algoritmos são empregados na concessão de crédito, recrutamento e seleção de profissionais, definição de benefícios sociais, diagnósticos médicos, seguros, educação e até mesmo em atividades relacionadas ao sistema de justiça.

Embora essas soluções possam aumentar a eficiência administrativa e reduzir custos operacionais, diversos estudos demonstram que decisões automatizadas não são necessariamente imparciais. Modelos treinados com bases de dados historicamente enviesadas tendem a reproduzir padrões discriminatórios já existentes, atribuindo aparência de neutralidade a decisões que podem reforçar desigualdades sociais.

O'Neil (2016) destaca que algoritmos de larga escala frequentemente influenciam oportunidades educacionais, profissionais e financeiras sem que seus critérios sejam transparentes ou passíveis de contestação pelos indivíduos afetados. Esse cenário torna ainda mais relevante a adoção de mecanismos de explicabilidade, auditoria e revisão humana, especialmente quando decisões automatizadas produzem impactos significativos sobre direitos individuais.

Outro desafio consiste na limitada capacidade de compreender como determinados modelos chegam às suas conclusões. Sistemas baseados em aprendizado profundo frequentemente apresentam elevado grau de complexidade, dificultando a interpretação de seu processo decisório. Conforme observam Krook et al. (2025), a ausência de transparência reduz a confiança nos sistemas inteligentes, dificulta auditorias independentes e compromete a efetividade dos mecanismos de responsabilização previstos pelos marcos regulatórios.

Nesse contexto, cresce a compreensão de que decisões automatizadas devem permanecer submetidas à supervisão humana, especialmente quando envolvem direitos fundamentais. Essa orientação está presente tanto na literatura especializada quanto em diversos instrumentos internacionais de governança da Inteligência Artificial.

2.3.3 Sistemas de armas autônomas

Entre as aplicações mais controversas da Inteligência Artificial encontram-se os Sistemas de Armas Autônomas Letais (*Lethal Autonomous Weapon Systems – LAWS*), capazes de identificar, selecionar e atacar alvos com diferentes níveis de autonomia operacional. O desenvolvimento dessas tecnologias tem provocado intenso debate envolvendo segurança internacional, direito humanitário, responsabilidade jurídica e ética militar.

Os principais questionamentos dizem respeito à possibilidade de delegar decisões potencialmente letais a sistemas computacionais. Diferentemente de armamentos convencionais, sistemas altamente autônomos podem reduzir significativamente a participação humana durante a identificação e o engajamento de alvos, tornando mais complexa a atribuição de responsabilidade em caso de falhas, erros operacionais ou violações do Direito Internacional Humanitário.

Bostrom (2014) e Russell (2019) argumentam que o desenvolvimento acelerado de sistemas inteligentes capazes de atuar em contextos militares exige mecanismos robustos de controle internacional, considerando os riscos decorrentes da corrida tecnológica entre Estados e da utilização de IA em conflitos armados. Da mesma forma, Noel Sharkey (2012) defende que decisões relacionadas ao uso da força letal não devem ser completamente delegadas a máquinas, uma vez que princípios como proporcionalidade, distinção e necessidade dependem de julgamentos éticos e jurídicos dificilmente reproduzíveis por sistemas automatizados.

Essas preocupações também são abordadas por Dafoe (2018), ao destacar que aplicações

militares da Inteligência Artificial podem alterar significativamente o equilíbrio estratégico entre Estados, ampliar riscos de escalada de conflitos e dificultar mecanismos internacionais de controle. Segundo o autor, a governança da IA deve contemplar não apenas aspectos tecnológicos, mas também seus impactos geopolíticos e humanitários.

A ausência de consenso internacional sobre a regulamentação dessas tecnologias evidencia a complexidade do tema. Embora organismos internacionais defendam a manutenção de controle humano significativo sobre decisões relacionadas ao emprego da força, ainda não existe um instrumento jurídico global capaz de disciplinar de forma uniforme o desenvolvimento e a utilização de sistemas de armas autônomas.

Em conjunto, os contextos analisados demonstram que os maiores desafios relacionados à Inteligência Artificial não decorrem exclusivamente da evolução tecnológica, mas principalmente dos impactos produzidos por sua aplicação em atividades capazes de restringir direitos, alterar relações sociais ou comprometer a segurança internacional. Essa constatação reforça a necessidade de mecanismos de governança que conciliem inovação tecnológica, gestão de riscos, supervisão humana e proteção dos direitos fundamentais, tema que será aprofundado na análise dos principais marcos regulatórios e diretrizes éticas aplicáveis à Inteligência Artificial.

3. METODOLOGIA

A presente pesquisa caracteriza-se como qualitativa, de natureza aplicada, com objetivos exploratórios e descritivos. A escolha dessa abordagem justifica-se pela necessidade de compreender como diferentes diretrizes éticas e instrumentos regulatórios tratam a governança da Inteligência Artificial, bem como identificar convergências, divergências e lacunas relacionadas à utilização dessas tecnologias em contextos considerados de alto risco.

Quanto aos procedimentos técnicos, o estudo fundamenta-se em pesquisa bibliográfica e documental. A pesquisa bibliográfica foi realizada a partir da análise de livros, artigos científicos e publicações nacionais e internacionais que abordam ética, governança e regulação da Inteligência Artificial. Foram consultadas obras de autores reconhecidos na área, como Stuart Russell, Nick Bostrom, Luciano Floridi, Allan Dafoe, Cathy O'Neil e Cary Coglianese, além de estudos publicados em periódicos científicos e conferências internacionais.

A pesquisa documental concentrou-se na análise de normas, diretrizes e instrumentos regulatórios relacionados à Inteligência Artificial, entre os quais se destacam o Regulamento Europeu de Inteligência Artificial (EU AI Act), o Regulamento Geral sobre a Proteção de Dados (GDPR), a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), os Princípios sobre Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), a Recomendação da UNESCO sobre a Ética da Inteligência Artificial e outros documentos institucionais considerados relevantes para a temática.

Como estratégia de investigação, adotou-se a análise comparativa dos principais referenciais éticos e regulatórios, buscando identificar os princípios comuns, as diferenças de abordagem e as lacunas existentes quanto à governança da Inteligência Artificial. A comparação foi direcionada, sobretudo, aos aspectos relacionados à transparência, responsabilização (*accountability*), equidade (*fairness*), explicabilidade (*explainability*), supervisão humana, gestão de riscos e proteção dos direitos fundamentais.

A análise concentrou-se em aplicações classificadas como de alto risco, considerando situações em que sistemas de Inteligência Artificial podem produzir impactos significativos sobre indivíduos ou coletividades. Entre esses contextos destacam-se aplicações voltadas à vigilância em massa, processos decisórios automatizados e sistemas de armas autônomas, escolhidos por representarem cenários frequentemente discutidos na literatura especializada e nos principais instrumentos regulatórios internacionais.

Os dados obtidos por meio da revisão bibliográfica e documental foram submetidos à análise

qualitativa, buscando interpretar as contribuições da literatura e dos marcos regulatórios para a construção de modelos de governança capazes de promover o desenvolvimento e a utilização responsável da Inteligência Artificial. A partir dessa análise, procurou-se identificar elementos que contribuam para o fortalecimento das práticas de governança em organizações públicas e privadas, considerando a necessidade de conciliar inovação tecnológica, conformidade regulatória e proteção dos direitos fundamentais.

A classificação metodológica adotada neste estudo está fundamentada nas contribuições de Gil (2019), Marconi e Lakatos (2021) e Creswell (2022), que compreendem a pesquisa qualitativa como instrumento adequado para a interpretação de fenômenos complexos e para a análise de documentos, normas e referenciais teóricos, especialmente em investigações voltadas à compreensão de processos sociais, jurídicos e organizacionais.

4. ANÁLISE CRÍTICA: CONVERGÊNCIAS E LACUNAS

A consolidação da Inteligência Artificial como tecnologia de amplo impacto social impulsionou a criação de diferentes instrumentos destinados a orientar seu desenvolvimento e utilização. Embora elaborados por organizações com competências e finalidades distintas, documentos como o EU AI Act, o GDPR, a Lei Geral de Proteção de Dados (LGPD), os Princípios de Inteligência Artificial da OCDE e a Recomendação da UNESCO sobre a Ética da Inteligência Artificial apresentam pontos de convergência relevantes, mas também diferenças quanto ao alcance, à obrigatoriedade e aos mecanismos de implementação.

Entre os aspectos convergentes, destaca-se o reconhecimento de que o desenvolvimento da Inteligência Artificial deve respeitar os direitos fundamentais, promover transparência nos processos decisórios, assegurar mecanismos de responsabilização e reduzir riscos de discriminação decorrentes do uso de algoritmos. Esses princípios aparecem, com diferentes níveis de detalhamento, em praticamente todos os referenciais analisados, evidenciando a existência de um núcleo comum de valores voltado à promoção de sistemas confiáveis e socialmente responsáveis.

Outro ponto de alinhamento refere-se à adoção da gestão de riscos como elemento central da governança da IA. O EU AI Act incorpora esse princípio ao classificar sistemas conforme seu potencial de impacto sobre direitos fundamentais e segurança, enquanto a OCDE e a UNESCO recomendam que organizações adotem medidas proporcionais aos riscos envolvidos em cada aplicação. Ainda que utilizem abordagens distintas, os documentos convergem ao reconhecer que sistemas de maior impacto exigem controles mais rigorosos, supervisão humana e monitoramento contínuo.

Apesar dessas convergências, permanecem lacunas importantes. A primeira delas diz respeito aos mecanismos de **enforcement**. Enquanto o EU AI Act e o GDPR estabelecem obrigações jurídicas, competências de fiscalização e sanções em caso de descumprimento, documentos como os Princípios da OCDE e a Recomendação da UNESCO possuem natureza orientativa. Embora exerçam influência sobre políticas públicas e boas práticas organizacionais, sua eficácia depende da adesão voluntária dos Estados e das organizações, inexistindo mecanismos internacionais capazes de impor seu cumprimento.

Essa diferença evidencia outra distinção relevante: a coexistência de instrumentos de **soft law** e **hard law** na regulação da Inteligência Artificial. O *soft law* compreende normas, recomendações e diretrizes que orientam comportamentos sem produzir efeitos jurídicos obrigatórios. Já o *hard law* corresponde às normas legalmente vinculantes, cuja observância pode ser exigida pelo Estado mediante mecanismos de fiscalização e aplicação de sanções. No contexto da IA, verifica-se que boa parte das iniciativas internacionais ainda se concentra em instrumentos de *soft law*, enquanto poucos ordenamentos jurídicos avançaram para modelos regulatórios obrigatórios.

Outra limitação observada refere-se ao caráter predominantemente principiológico das diretrizes éticas. Embora documentos internacionais estabeleçam valores como transparência,

equidade e responsabilização, raramente definem critérios objetivos para avaliar seu cumprimento ou indicadores que permitam medir sua efetividade. Essa característica dificulta a implementação uniforme desses princípios pelas organizações e amplia a possibilidade de interpretações distintas sobre seu significado e aplicação prática.

Também merecem destaque as diferenças culturais e institucionais existentes entre os países. A concepção de privacidade, liberdade de expressão, segurança pública e proteção de dados varia conforme fatores históricos, sociais e políticos. Como consequência, modelos regulatórios desenvolvidos em determinado contexto nem sempre podem ser reproduzidos integralmente em outras realidades. A abordagem adotada pela União Europeia, fortemente orientada pela proteção dos direitos fundamentais, difere daquela observada em países que priorizam o desenvolvimento tecnológico, a segurança nacional ou a competitividade econômica. Essa diversidade demonstra que a construção de uma governança global da Inteligência Artificial enfrenta desafios que extrapolam aspectos técnicos e jurídicos.

Outro aspecto identificado na literatura refere-se à velocidade da evolução tecnológica. Enquanto sistemas de Inteligência Artificial são constantemente atualizados, os processos legislativos tendem a ocorrer em ritmo mais lento. Essa assimetria dificulta a elaboração de normas capazes de acompanhar a rápida transformação das tecnologias, tornando necessária a adoção de modelos regulatórios suficientemente flexíveis para responder a novos riscos sem comprometer a inovação.

A análise realizada permite concluir que existe crescente convergência em torno dos princípios que devem orientar o desenvolvimento responsável da Inteligência Artificial. Entretanto, ainda persistem diferenças relevantes quanto aos mecanismos de implementação, ao grau de obrigatoriedade das normas e à capacidade de fiscalização dos instrumentos existentes. Nesse cenário, observa-se que a efetividade da governança da IA depende não apenas da existência de princípios éticos ou de normas jurídicas, mas também da capacidade das organizações de incorporá-los em suas estruturas de decisão, processos internos e práticas de gestão de riscos.

Tabela 1

Instrumento	Natureza	Transparência	Accountability	Fairness	Explicabilidade	Gestão de Riscos
LGPD (Brasil)	Hard Law	☐	☐	Parcial	Parcial	Parcial
GDPR (União Europeia)	Hard Law	☐	☐	☐	☐	Parcial
EU AI Act (União Europeia)	Hard Law	☐	☐	☐	☐	☐
Princípios da OCDE	Soft Law	☐	☐	☐	Parcial	☐
Recomendação da UNESCO	Soft Law	☐	☐	☐	☐	☐

Fonte: o autor

Nota: "Parcial" indica que o princípio está presente de forma indireta ou não constitui obrigação expressa no respectivo instrumento.

A Tabela 1 evidencia que, embora os principais instrumentos analisados compartilhem valores como transparência, responsabilização e mitigação de riscos, existem diferenças significativas quanto à natureza jurídica e ao grau de detalhamento das exigências. Enquanto o EU AI Act estabelece obrigações específicas para sistemas de alto risco, os princípios da OCDE e a Recomendação da UNESCO possuem caráter orientativo, funcionando como referenciais para a formulação de

5. IMPACTOS PARA AS ORGANIZAÇÕES

A adoção crescente de sistemas de Inteligência Artificial em atividades estratégicas tem ampliado os benefícios relacionados à eficiência operacional, à automação de processos e ao suporte à tomada de decisão. Entretanto, quando esses sistemas são desenvolvidos ou utilizados sem estruturas adequadas de governança, as organizações passam a enfrentar riscos que transcendem aspectos tecnológicos, alcançando dimensões jurídicas, reputacionais e até institucionais. Nesse contexto, a governança da IA deixa de representar apenas uma boa prática de gestão para tornar-se um elemento essencial da estratégia organizacional.

5.1 Riscos jurídicos

Os riscos jurídicos decorrentes do uso inadequado da Inteligência Artificial estão diretamente relacionados ao descumprimento de normas de proteção de dados, direitos fundamentais, responsabilidade civil e exigências regulatórias específicas. Em diferentes jurisdições, observa-se uma tendência de fortalecimento dos mecanismos de responsabilização das organizações que utilizam sistemas de IA, especialmente quando estes produzem decisões capazes de afetar indivíduos.

No contexto brasileiro, a Lei Geral de Proteção de Dados (LGPD) estabelece princípios como finalidade, necessidade, transparência e não discriminação, além de assegurar ao titular o direito de solicitar revisão de decisões tomadas exclusivamente por tratamento automatizado de dados pessoais. Embora a LGPD não constitua uma legislação específica sobre Inteligência Artificial, diversos de seus dispositivos tornam-se diretamente aplicáveis aos sistemas inteligentes que realizam tratamento de dados pessoais.

Em âmbito internacional, o Regulamento Geral sobre a Proteção de Dados (GDPR) e o Regulamento Europeu de Inteligência Artificial (EU AI Act) ampliam esse cenário ao estabelecer obrigações específicas para sistemas classificados como de alto risco. Entre essas exigências destacam-se a realização de avaliações de risco, documentação técnica, rastreabilidade, supervisão humana, monitoramento contínuo e mecanismos de gestão da qualidade. O descumprimento dessas obrigações pode resultar em sanções administrativas, responsabilização civil e restrições ao uso da tecnologia.

Além das penalidades legais, organizações podem ser responsabilizadas por danos decorrentes de decisões algorítmicas discriminatórias, falhas de segurança, uso indevido de dados pessoais ou ausência de mecanismos adequados de supervisão humana. Dessa forma, a conformidade regulatória passa a representar não apenas uma obrigação legal, mas também um instrumento de mitigação de riscos organizacionais.

5.2 Riscos reputacionais

Os impactos reputacionais decorrentes do uso inadequado da Inteligência Artificial podem ser tão significativos quanto as consequências jurídicas. A confiança constitui um ativo estratégico para organizações públicas e privadas, sendo diretamente influenciada pela percepção de transparência, responsabilidade e respeito aos direitos dos usuários.

Incidentes envolvendo vazamento de dados, discriminação algorítmica, decisões automatizadas injustificadas ou utilização de sistemas opacos tendem a receber ampla repercussão na mídia e nas redes sociais, comprometendo a imagem institucional. Em mercados altamente competitivos, a perda de confiança pode afetar relacionamentos com clientes, investidores, parceiros comerciais e órgãos reguladores.

Heilinger (2022) observa que a simples adoção de princípios éticos não é suficiente para assegurar legitimidade institucional. Segundo o autor, a utilização de códigos éticos apenas como estratégia de marketing ou para evitar regulamentações mais rigorosas — prática conhecida como

ethics washing — pode produzir efeito contrário ao esperado, ampliando a desconfiança da sociedade e comprometendo a credibilidade das organizações.

Nesse cenário, políticas efetivas de governança da IA contribuem para fortalecer a reputação institucional ao demonstrar compromisso com transparência, prestação de contas, gestão de riscos e respeito aos direitos fundamentais.

5.3 Falhas de governança

Grande parte dos problemas associados à Inteligência Artificial decorre menos das limitações tecnológicas e mais da ausência de mecanismos organizacionais capazes de orientar seu desenvolvimento e utilização. Falhas de governança costumam envolver inexistência de políticas internas, indefinição de responsabilidades, ausência de avaliações de impacto, monitoramento insuficiente, deficiência na documentação dos modelos e inexistência de processos de auditoria.

Outro problema recorrente consiste na utilização de modelos treinados com bases de dados inadequadas ou desatualizadas, sem mecanismos de validação contínua. Nessas circunstâncias, vieses presentes nos dados podem ser reproduzidos ou ampliados pelos algoritmos, comprometendo a qualidade das decisões e expondo a organização a riscos éticos e legais.

A literatura recente destaca ainda que modelos de governança eficazes devem contemplar todo o ciclo de vida dos sistemas de IA, desde sua concepção até a descontinuação. Isso inclui gestão contínua de riscos, documentação das decisões, monitoramento de desempenho, atualização dos modelos e supervisão humana permanente (Coglianese; Crum, 2025). Sob essa perspectiva, a governança da IA aproxima-se das práticas modernas de gestão de riscos corporativos, incorporando controles preventivos e mecanismos permanentes de melhoria contínua.

5.4 Casos práticos

Diversos casos amplamente documentados evidenciam as consequências da ausência de práticas adequadas de governança da Inteligência Artificial.

Um dos exemplos mais conhecidos ocorreu na Amazon, cujo sistema experimental de recrutamento automatizado demonstrou tendência a desfavorecer candidatas mulheres durante os processos seletivos. O algoritmo havia sido treinado com dados históricos predominantemente masculinos, reproduzindo padrões discriminatórios existentes na base de treinamento. Diante da impossibilidade de eliminar satisfatoriamente o viés identificado, o projeto foi descontinuado (Dastin, 2018).

Outro caso emblemático refere-se ao sistema COMPAS, utilizado nos Estados Unidos para estimar risco de reincidência criminal. Investigações conduzidas pela organização ProPublica indicaram que o modelo apresentava diferenças significativas entre grupos raciais, classificando indivíduos negros como de maior risco em proporção superior à observada para indivíduos brancos em situações equivalentes. Embora existam debates metodológicos sobre os critérios estatísticos empregados, o caso tornou-se referência internacional nas discussões sobre vieses algorítmicos e transparência em sistemas de apoio à decisão judicial.

Também merece destaque o episódio envolvendo reconhecimento facial utilizado por autoridades policiais em diferentes países. Estudos conduzidos por Buolamwini e Gebru (2018) demonstraram que diversos sistemas comerciais apresentavam taxas de erro superiores na identificação de mulheres e pessoas negras quando comparadas às observadas para homens brancos. Esses resultados impulsionaram revisões regulatórias, restrições ao uso dessa tecnologia em determinados contextos e maior exigência de auditorias independentes para aplicações de alto risco.

Esses exemplos evidenciam que os principais riscos associados à Inteligência Artificial não decorrem exclusivamente da tecnologia, mas da ausência de governança capaz de assegurar qualidade dos dados, transparência, supervisão humana, prestação de contas e monitoramento contínuo. Assim, organizações que incorporam práticas estruturadas de governança reduzem sua exposição a riscos

jurídicos e reputacionais, fortalecem a confiança dos diferentes stakeholders e aumentam a legitimidade de suas iniciativas baseadas em Inteligência Artificial.

6. RECOMENDAÇÕES PARA A GOVERNANÇA DA INTELIGÊNCIA ARTIFICIAL

O avanço da Inteligência Artificial nas organizações exige que sua implementação seja acompanhada por mecanismos capazes de assegurar conformidade legal, responsabilidade ética e efetiva gestão de riscos. Nesse contexto, a governança de IA deve ser compreendida como um modelo organizacional integrado, que articula princípios éticos, compliance, gestão de riscos e supervisão contínua ao longo de todo o ciclo de vida dos sistemas inteligentes. Essa abordagem reduz a exposição a riscos regulatórios e reputacionais, ao mesmo tempo em que fortalece a confiança de usuários, investidores e órgãos reguladores (OECD, 2024; NIST, 2023).

6.1 Alinhamento entre ética e compliance

Embora ética e compliance possuam finalidades distintas, ambas devem atuar de forma complementar na governança da Inteligência Artificial. Enquanto o compliance busca assegurar o cumprimento das obrigações legais, regulatórias e normativas, a ética amplia essa perspectiva ao incorporar valores relacionados à justiça, transparência, responsabilidade, inclusão e respeito aos direitos fundamentais.

Nesse sentido, limitar a governança da IA ao atendimento das exigências legais revela-se insuficiente. Heilinger (2022) argumenta que uma governança efetivamente ética deve considerar também os impactos sociais, econômicos e políticos produzidos pelos sistemas inteligentes, evitando uma visão restrita centrada apenas na conformidade normativa. Segundo o autor, a participação de especialistas independentes, representantes da sociedade e diferentes partes interessadas contribui para decisões mais equilibradas e legítimas, reduzindo o risco de práticas conhecidas como *ethics washing*, nas quais princípios éticos são utilizados apenas como instrumento de marketing institucional, sem efetiva aplicação prática.

Sob essa perspectiva, programas de compliance devem incorporar procedimentos específicos para sistemas de IA, incluindo avaliações de impacto, critérios para classificação de riscos, controles sobre qualidade dos dados, supervisão humana, rastreabilidade das decisões automatizadas e mecanismos para tratamento de incidentes. A integração entre compliance, governança corporativa e gestão de riscos permite que aspectos éticos sejam considerados desde as etapas iniciais de desenvolvimento das soluções, seguindo o princípio do *ethics by design*.

6.2 Frameworks práticos para governança de IA

Nos últimos anos, diferentes organismos internacionais desenvolveram estruturas de referência destinadas a orientar organizações na implementação de sistemas de IA confiáveis e responsáveis.

Entre os principais modelos destaca-se o NIST AI Risk Management Framework (AI RMF 1.0), publicado pelo National Institute of Standards and Technology dos Estados Unidos. O framework propõe uma abordagem contínua de gestão de riscos baseada em quatro funções interdependentes: Govern, Map, Measure e Manage. O modelo enfatiza que os riscos associados à IA devem ser monitorados continuamente, considerando aspectos técnicos, organizacionais, sociais e jurídicos ao longo de todo o ciclo de vida dos sistemas.

Outro importante referencial é o conjunto de Princípios para Inteligência Artificial da Organização para Cooperação e Desenvolvimento Econômico (OCDE), adotado por dezenas de países. Esses princípios estabelecem que sistemas de IA devem promover crescimento inclusivo, respeitar direitos humanos, assegurar transparência, robustez técnica, segurança, prestação de contas e supervisão humana apropriada.

No contexto europeu, o EU AI Act consolida uma abordagem regulatória baseada em níveis

de risco, classificando aplicações de IA em categorias que variam de risco mínimo até risco inaceitável. Para sistemas considerados de alto risco, a legislação exige requisitos como gestão formal de riscos, documentação técnica, governança de dados, monitoramento contínuo, transparência, supervisão humana e registro das operações realizadas.

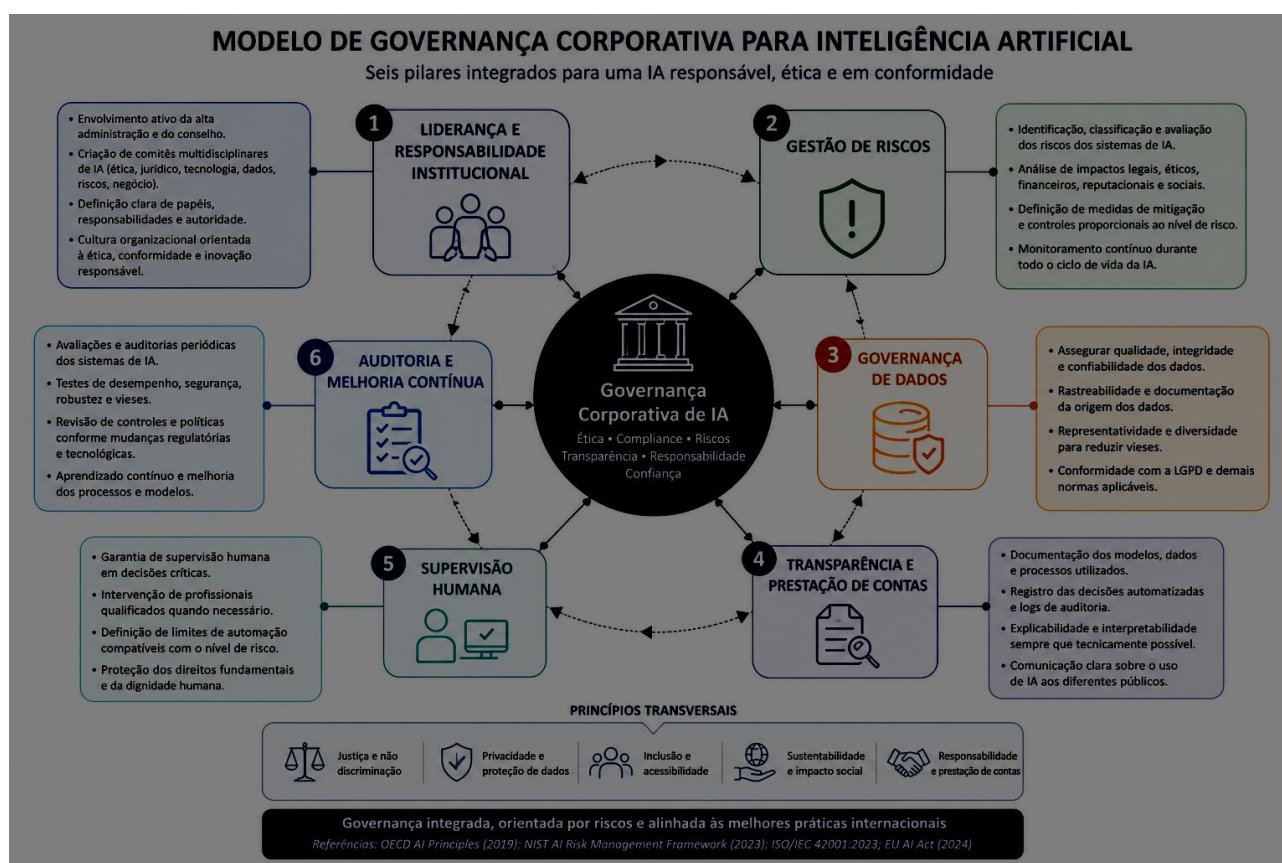
Além desses referenciais, a norma ISO/IEC 42001:2023, primeiro padrão internacional para Sistemas de Gestão de Inteligência Artificial, oferece uma estrutura semelhante aos sistemas de gestão já consolidados, como as normas ISO 27001 (segurança da informação) e ISO 37301 (compliance). Seu objetivo é integrar políticas, processos, controles e melhoria contínua à governança organizacional da IA, favorecendo maior maturidade institucional.

A convergência entre esses frameworks evidencia um consenso internacional de que a governança da IA deve ultrapassar aspectos exclusivamente tecnológicos, incorporando gestão de riscos, responsabilidade organizacional, transparência e monitoramento permanente.

6.3 Proposta de governança corporativa para Inteligência Artificial

Considerando as recomendações presentes no NIST AI Risk Management Framework (2023), na norma ISO/IEC 42001:2023, nos Princípios de IA da OCDE (OECD, 2019) e no Regulamento Europeu de Inteligência Artificial (EU AI Act), propõe-se um modelo de governança corporativa estruturado em seis pilares complementares: (i) liderança e responsabilidade institucional; (ii) gestão de riscos; (iii) governança de dados; (iv) transparência e prestação de contas; (v) supervisão humana; e (vi) auditoria e melhoria contínua.

Figura 2 – Modelo proposto de Governança Corporativa para Inteligência Artificial



Fonte: Elaborado pelo autor (2026), com base em OECD (2019), HLEG (2019), NIST AI RMF 1.0 (2023), ISO/IEC 42001:2023 e EU AI Act (2024)

O primeiro pilar corresponde à liderança e responsabilidade institucional, com definição clara das atribuições da alta administração, criação de comitês multidisciplinares de IA e estabelecimento de papéis específicos para supervisão dos sistemas inteligentes.

O segundo pilar refere-se à gestão de riscos, envolvendo identificação, classificação, avaliação

e tratamento dos riscos associados aos modelos de IA antes de sua implantação, bem como monitoramento contínuo durante sua operação.

O terceiro pilar contempla a governança de dados, assegurando qualidade, integridade, rastreabilidade e representatividade das bases utilizadas no treinamento dos modelos, reduzindo a probabilidade de vieses e resultados discriminatórios.

O quarto pilar concentra-se na transparência e prestação de contas, mediante documentação dos modelos, registro das decisões automatizadas, explicabilidade sempre que tecnicamente possível e comunicação clara sobre o uso da Inteligência Artificial aos diferentes públicos envolvidos.

O quinto pilar envolve a supervisão humana, garantindo que decisões críticas permaneçam sujeitas à intervenção de profissionais qualificados, especialmente em aplicações que afetem direitos fundamentais ou produzam impactos sobre indivíduos.

Por fim, o sexto pilar corresponde à auditoria e melhoria contínua, compreendendo avaliações periódicas dos sistemas, testes de desempenho, revisão de controles, atualização dos modelos e incorporação das mudanças regulatórias e tecnológicas.

Esse modelo está alinhado às recomendações da OCDE, do NIST, da ISO/IEC 42001 e da regulamentação europeia, permitindo que organizações desenvolvam uma governança integrada, orientada por riscos e compatível com as exigências regulatórias emergentes.

Conforme destaca Heilinger (2022), uma governança efetivamente ética não depende apenas da existência de normas ou princípios abstratos, mas da inclusão de mecanismos institucionais capazes de assegurar participação, responsabilização e avaliação crítica dos impactos produzidos pela Inteligência Artificial. Para o autor, soluções exclusivamente tecnológicas são insuficientes para enfrentar problemas que possuem natureza social e institucional, sendo necessária uma governança que considere também justiça, participação dos atores afetados e responsabilidade organizacional ao longo de todo o ciclo de vida da tecnologia.

Assim, recomenda-se que as organizações adotem uma abordagem integrada, na qual ética, compliance, gestão de riscos e governança corporativa atuem de forma coordenada. Mais do que atender exigências regulatórias, esse modelo contribui para o desenvolvimento de sistemas de Inteligência Artificial confiáveis, transparentes e alinhados aos princípios de responsabilidade, sustentabilidade e respeito aos direitos fundamentais, fortalecendo a legitimidade institucional e a confiança dos diversos stakeholders.

7. CONCLUSÃO

A análise realizada ao longo deste estudo demonstra que a expansão da Inteligência Artificial tem sido acompanhada pela criação de diferentes diretrizes éticas e instrumentos regulatórios destinados a orientar seu desenvolvimento e utilização responsável. Apesar das diferenças de natureza e alcance, os referenciais analisados apresentam convergências importantes em torno de princípios como transparência, responsabilização, equidade, proteção dos direitos fundamentais, supervisão humana e gestão de riscos. A proposta de pesquisa partiu justamente da identificação de possíveis lacunas entre esses princípios e sua efetiva incorporação às práticas organizacionais.

Em relação à primeira questão de pesquisa, verificou-se que diretrizes éticas e marcos regulatórios compartilham um núcleo relevante de princípios voltados ao uso responsável da IA, mas diferem quanto ao grau de obrigatoriedade e aos mecanismos disponíveis para sua implementação e fiscalização. Enquanto instrumentos de *hard law*, como a LGPD, o GDPR e o EU AI Act, estabelecem obrigações jurídicas e mecanismos de responsabilização, referenciais como os Princípios da OCDE e a Recomendação da UNESCO possuem caráter predominantemente orientativo. Dessa forma, a existência de princípios éticos amplamente reconhecidos não garante, por si só, sua aplicação efetiva nas organizações.

Quanto à segunda questão, a análise dos contextos sensíveis evidenciou que aplicações

relacionadas à vigilância, decisões automatizadas e sistemas de armas autônomas demandam níveis mais rigorosos de controle. Nessas situações, eventuais falhas podem ultrapassar prejuízos operacionais e atingir diretamente direitos fundamentais, a segurança e a dignidade das pessoas. O problema identificado, portanto, não está apenas na capacidade técnica dos sistemas, mas também na forma como são utilizados, supervisionados e submetidos a mecanismos de responsabilização.

A terceira questão, relacionada à possibilidade de adoção de um padrão global para as normas éticas de IA, apresenta uma resposta menos objetiva. A pesquisa indica que a existência de princípios comuns é desejável, sobretudo no que diz respeito à proteção dos direitos fundamentais, à transparência e à responsabilização. Entretanto, sua aplicação não pode desconsiderar as diferenças jurídicas, culturais, sociais e institucionais existentes entre os países. Nesse sentido, a harmonização de princípios parece mais viável do que a adoção de um modelo regulatório único e uniforme.

Os resultados também reforçam a importância de integrar ética e governança às estruturas organizacionais. A literatura analisada demonstra que a ética em IA não deve permanecer restrita a declarações de princípios, mas precisa ser incorporada aos processos de desenvolvimento, implementação e monitoramento dos sistemas. Heilinger (2022), por exemplo, chama atenção para a necessidade de considerar as condições sociais, econômicas e políticas que envolvem o desenvolvimento e a utilização dessas tecnologias, ampliando a análise ética para além do funcionamento específico de cada aplicação.

Da mesma forma, estudos recentes sobre transparência indicam que sua adoção não deve ser tratada como solução isolada para os problemas de governança. A literatura aponta a necessidade de uma abordagem que envolva toda a organização, incorporando a transparência à cultura institucional e às diferentes etapas do desenvolvimento e implementação dos sistemas de IA.

Conclui-se, portanto, que a governança responsável da Inteligência Artificial depende da articulação entre princípios éticos, exigências regulatórias e mecanismos organizacionais capazes de transformar esses referenciais em práticas concretas. A simples existência de normas ou diretrizes não elimina os riscos associados à IA. É necessário que organizações estabeleçam responsabilidades claras, processos de avaliação e gestão de riscos, mecanismos de supervisão e formas adequadas de monitoramento e prestação de contas. A negligência desses aspectos pode ampliar a exposição a riscos jurídicos, reputacionais e sociais, especialmente em aplicações de alto impacto.

Como limitação, destaca-se que o presente estudo foi desenvolvido a partir de pesquisa bibliográfica e documental, não contemplando investigação empírica junto a organizações ou profissionais responsáveis pela implementação de sistemas de IA. Dessa forma, os resultados refletem a análise dos referenciais selecionados e da literatura consultada, não permitindo generalizações sobre o nível efetivo de maturidade das organizações.

Para pesquisas futuras, recomenda-se a realização de estudos empíricos que investiguem como organizações brasileiras estão estruturando suas práticas de governança de IA e em que medida princípios éticos e requisitos regulatórios são efetivamente incorporados aos processos internos. Também se mostra pertinente investigar a aplicação prática de frameworks como o NIST AI RMF e a ISO/IEC 42001, bem como comparar experiências de diferentes setores e países. Estudos futuros poderão, ainda, aprofundar a relação entre proteção de dados, segurança da informação, gestão de riscos e governança de sistemas de IA em contextos considerados de alto risco.

REFERÊNCIAS

BOSTROM, Nick. *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press, 2014.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 01 jul. 2026.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in

commercial gender classification. *Proceedings of Machine Learning Research*, v. 81, p. 77-91, 2018.

COGLIANESE, Cary; CRUM, Colton R. Leashes, not guardrails: a management-based approach to artificial intelligence risk regulation. *Risk Analysis*, v. 45, n. 12, p. 4397-4407, 2025. DOI: 10.1111/risa.70020.

CRESWELL, John W.; CRESWELL, J. David. *Research design: qualitative, quantitative, and mixed methods approaches*. 6. ed. Thousand Oaks: SAGE Publications, 2022.

DAFOE, Allan. *AI governance: a research agenda*. Oxford: Centre for the Governance of AI, Future of Humanity Institute, University of Oxford, 2018.

DASTIN, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10 out. 2018. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>. Acesso em: 25 jul. 2026.

EUROPEAN PARLIAMENT; COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016: General Data Protection Regulation (GDPR). *Official Journal of the European Union*, 2016. Disponível em: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Acesso em: 26 jul. 2026.

EUROPEAN PARLIAMENT; COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024: laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, 2024. Disponível em: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>. Acesso em: 26 jul.2026.

FLORIDI, Luciano; COWLS, Josh. A unified framework of five principles for AI in society. *Harvard Data Science Review*, v. 1, n. 1, 2019. DOI: 10.1162/99608f92.8cd550d1.

GIL, Antonio Carlos. *Métodos e técnicas de pesquisa social*. 7. ed. São Paulo: Atlas, 2019.

HEILINGER, Jan-Christoph. The ethics of AI ethics: a constructive critique. *Philosophy & Technology*, v. 35, art. 61, 2022. DOI: 10.1007/s13347-022-00557-9.

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. *Ethics guidelines for trustworthy AI*. Brussels: European Commission, 2019. DOI: 10.2759/346720.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; INTERNATIONAL ELECTROTECHNICAL COMMISSION. *ISO/IEC 42001:2023: information technology — artificial intelligence — management system*. Geneva: ISO, 2023.

KIM, Byeong-Je et al. AI governance in the context of the EU AI Act: a bibliometric and literature review approach. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, v. 8, n. 2, p. 1424-1425, 2025. DOI: 10.1609/aies.v8i2.36642.

KROOK, Joshua et al. A systematic literature review of artificial intelligence (AI) transparency laws in the European Union (EU) and United Kingdom (UK): a socio-legal approach to AI transparency governance. *AI and Ethics*, v. 5, n. 4, p. 4069-4090, 2025. DOI: 10.1007/s43681-025-00674-z.

MÄNTYMÄKI, Matti et al. Defining organizational AI governance. *AI and Ethics*, v. 2, p. 603-609, 2022. DOI: 10.1007/s43681-022-00143-x.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Metodologia científica*. 9. ed. São Paulo: Atlas, 2021.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). *Artificial intelligence risk management framework (AI RMF 1.0)*. Gaithersburg: National Institute of Standards and Technology, 2023. NIST AI 100-1. DOI: 10.6028/NIST.AI.100-1.

O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishers, 2016.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Recommendation of the Council on Artificial Intelligence*. Paris: OECD, 2019. OECD/LEGAL/0449.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Recommendation of the Council on Artificial Intelligence: 2024 update*. Paris: OECD, 2024.

PAHUNE, Saurabh et al. The importance of AI data governance in large language models. *Big Data and Cognitive Computing*, v. 9, n. 6, art. 147, 2025. DOI: 10.3390/bdcc9060147.

RUSSELL, Stuart. *Human compatible: artificial intelligence and the problem of control*. New York: Viking, 2019.

RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 4. ed. Hoboken: Pearson, 2021.

SHARKEY, Noel. Mindless killer robots must be stopped. *The Guardian*, 3 dez. 2012. Disponível em: <https://www.theguardian.com/commentisfree/2012/dec/03/mindless-killer-robots>. Acesso em: 27 jul.2026.

UNESCO. *Recommendation on the ethics of artificial intelligence*. Paris: UNESCO, 2022. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Acesso em: 30 jul.2026.