

Autoavaliação de maturidade no tratamento de riscos de IA**Maturity self-assessment in AI risk treatment****Autoevaluación de madurez en el tratamiento de riesgos de IA****Goudim Alvarenga Carneiro**

Especialista em Gestão Pública

Instituição: Escola Nacional de Administração Pública (ENAP)

Endereço: Brasília – Distrito Federal, Brasil

E-mail: goudim@gmail.com

João Souza Neto

Pós-Doutor em Segurança Cibernética

Instituição: Universidade de Brasília (UNB)

Endereço: Brasília – Distrito Federal, Brasil

E-mail: sznetoj@gmail.com

RESUMO

A crescente adoção de sistemas de inteligência artificial (IA) por organizações públicas e privadas impõe desafios relacionados à governança e à avaliação de riscos, frequentemente sem estruturas institucionais consolidadas para enfrentá-los. Apesar da publicação de referenciais normativos internacionais relevantes, em especial a ISO/IEC 42001, que estabelece requisitos para sistemas de gestão de IA, e a ISO/IEC 23894, que orienta a gestão de riscos de IA, observa-se a ausência de instrumentos integrados capazes de articular sistematicamente a avaliação da maturidade dos controles com a exposição às fontes de risco identificadas nessas normas. Este estudo teve como objetivo desenvolver um instrumento integrado de autoavaliação organizacional que articula os controles do Anexo A da ISO/IEC 42001 com as fontes de risco do Anexo B da ISO/IEC 23894, por meio de um mapeamento normativo rastreável. A pesquisa é de natureza aplicada, com abordagem quantitativa e caráter exploratório, adotando um raciocínio dedutivo a partir dos referenciais normativos. O instrumento foi submetido a uma aplicação piloto em nove organizações de perfis distintos, públicas e privadas, de diferentes portes, setores e estágios de adoção de IA. Os resultados indicam que o instrumento é operacionalmente aplicável em contextos heterogêneos, que as escalas utilizadas discriminam adequadamente diferentes percepções organizacionais, e que o cruzamento entre os blocos de avaliação gera saídas diagnósticas internamente coerentes com a lógica do mapeamento proposto. Conclui-se que o instrumento oferece uma contribuição metodológica ao campo da governança de IA, ao propor um método estruturado, normativamente fundamentado e acessível ao autodiagnóstico organizacional.

Palavras-chave: governança de Inteligência Artificial, gestão de riscos, maturidade organizacional, ISO/IEC 42001, ISO/IEC 23894.

ABSTRACT

The growing adoption of artificial intelligence (AI) systems by public and private organizations poses challenges related to governance and risk assessment, often in the absence of consolidated

institutional structures to address them. Despite the publication of relevant international normative frameworks, particularly ISO/IEC 42001, which establishes requirements for AI management systems, and ISO/IEC 23894, which provides guidance on AI risk management, there is a notable lack of integrated instruments capable of systematically articulating the assessment of control maturity with exposure to the risk sources identified in these standards. This study aimed to develop an integrated organizational self-assessment instrument that maps the controls in Annex A of ISO/IEC 42001 to the risk sources in Annex B of ISO/IEC 23894 through a traceable normative mapping. The research is applied in nature, adopts a quantitative approach with exploratory scope, and follows a deductive method grounded in the normative frameworks. The instrument was piloted in nine organizations with distinct profiles, public and private, of different sizes, sectors, and AI adoption stages. Results indicate that the instrument is operationally applicable across heterogeneous contexts, that the scales used adequately discriminate among organizational perceptions, and that cross-analysis of the assessment blocks yields diagnostic outputs that are internally consistent with the logic of the proposed mapping. The study concludes that the instrument makes a methodological contribution to the field of AI governance by proposing a structured, normatively grounded, and accessible method for organizational self-diagnosis.

Keywords: Artificial Intelligence governance, risk management, organizational maturity, ISO/IEC 42001, ISO/IEC 23894.

RESUMEN

La creciente adopción de sistemas de inteligencia artificial (IA) por parte de organizaciones públicas y privadas plantea desafíos en materia de gobernanza y evaluación de riesgos, a menudo sin estructuras institucionales consolidadas para afrontarlos. A pesar de la publicación de marcos normativos internacionales relevantes, en especial la ISO/IEC 42001, que establece requisitos para los sistemas de gestión de IA, y la ISO/IEC 23894, que orienta la gestión de riesgos de IA, se observa la ausencia de instrumentos integrados capaces de articular sistemáticamente la evaluación de la madurez de los controles con la exposición a las fuentes de riesgo identificadas en dichas normas. Este estudio tuvo como objetivo desarrollar un instrumento integrado de autoevaluación organizacional que articula los controles del Anexo A de la ISO/IEC 42001 con las fuentes de riesgo del Anexo B de la ISO/IEC 23894, mediante un mapeo normativo trazable. La investigación es de naturaleza aplicada, adopta un enfoque cuantitativo de carácter exploratorio y sigue un método deductivo basado en los marcos normativos. El instrumento fue sometido a una aplicación piloto en nueve organizaciones con perfiles distintos, públicas y privadas, de diferentes tamaños, sectores y etapas de adopción de IA. Los resultados indican que el instrumento es operacionalmente aplicable en contextos heterogéneos, que las escalas utilizadas discriminan adecuadamente distintas percepciones organizacionales, y que el cruce entre los bloques de evaluación genera salidas diagnósticas internamente coherentes con la lógica del mapeo propuesto. Se concluye que el instrumento aporta una contribución metodológica al campo de la gobernanza de IA, al proponer un método estructurado, normativamente fundamentado y accesible para el autodiagnóstico organizacional.

Palabras clave: gobernanza de Inteligencia Artificial, gestión de riesgos, madurez organizacional, ISO/IEC 42001, ISO/IEC 23894.

1 INTRODUÇÃO

A incorporação de soluções baseadas em Inteligência Artificial (IA) tem se expandido significativamente nas últimas décadas, impulsionada pelo aumento da capacidade computacional, pela ampla disponibilidade de dados e pelo avanço das técnicas de aprendizado de máquina. Esse movimento vem transformando não apenas o setor privado, mas também as estruturas administrativas dos governos, que passaram a empregar sistemas de IA em atividades como a automação de processos, a análise de grandes volumes de dados, o apoio à tomada de decisão e a prestação de serviços digitais à população (OECD, 2019).

Diversos organismos internacionais têm desenvolvido diretrizes e referenciais para orientar o uso responsável da inteligência artificial. Entre essas iniciativas destacam-se os princípios da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) para uma IA confiável, que enfatizam aspectos como robustez técnica, segurança, responsabilidade e governança adequada (OECD, 2019). De forma complementar, o National Institute of Standards and Technology (NIST) publicou o AI Risk Management Framework (AI RMF), que propõe uma abordagem estruturada para a identificação, avaliação e mitigação de riscos associados ao desenvolvimento e à implementação de sistemas de IA nas organizações (NIST, 2023).

No campo regulatório, observa-se o avanço de iniciativas voltadas à criação de marcos legais específicos para inteligência artificial. Um dos exemplos mais relevantes é o Artificial Intelligence Act da União Europeia, que estabelece uma estrutura regulatória baseada em níveis de risco, definindo requisitos de transparência, governança de dados, avaliação de conformidade e supervisão institucional para aplicações consideradas de alto risco (European Union, 2024). Essas iniciativas refletem um movimento global de institucionalização de práticas voltadas ao desenvolvimento seguro, responsável e confiável de sistemas de IA.

Nesse contexto, a publicação da ISO/IEC 42001, em 2023, representa um avanço relevante ao estabelecer requisitos para a implementação de Sistemas de Gestão de Inteligência Artificial (Artificial Intelligence Management System, AIMS). A norma adota a estrutura de alto nível das normas de sistemas de gestão da família ISO, baseada no ciclo de melhoria contínua (Plan–Do–Check–Act), e organiza os requisitos relacionados ao contexto organizacional, liderança, planejamento, operação, avaliação de desempenho e melhoria contínua (ISO, 2023). Seu objetivo é apoiar as organizações na criação de políticas, processos e controles capazes de

assegurar o desenvolvimento e o uso responsável de sistemas de IA ao longo de todo o ciclo de vida desses sistemas.

No campo da inteligência artificial, essa discussão também se relaciona à distinção entre governança e gestão. A governança de IA estabelece os princípios, responsabilidades e mecanismos de accountability que orientam o desenvolvimento e a implantação dos sistemas, enquanto a gestão de IA corresponde à operacionalização dessas diretrizes ao longo do ciclo de vida da IA, por meio da execução de processos, controles e atividades de avaliação de riscos (Raji et al., 2020).

Em organizações de qualquer natureza, essa capacidade institucional torna-se cada vez mais relevante. Embora organizações, especialmente no setor público, tenham ampliado o uso de tecnologias de IA em áreas como regulação, prestação de serviços, tomada de decisão e otimização de processos, estudos indicam que essa adoção ocorre frequentemente de forma experimental ou fragmentada, em contextos em que práticas organizacionais ainda se mostram insuficientemente estruturadas. (Veale; Van Kleek; Binns, 2018).

Embora a literatura disponha de modelos de maturidade para IA, frameworks de gestão de riscos e normas para sistemas de gestão, observa-se uma integração limitada entre esses referenciais em instrumentos capazes de avaliar, de forma conjunta, a maturidade organizacional e a exposição a riscos em inteligência artificial. Em geral, os modelos existentes privilegiam dimensões específicas, como governança, IA responsável, auditoria algorítmica ou gestão de riscos, sem combinar, em um único instrumento, os requisitos de um sistema de gestão de IA com a avaliação estruturada das fontes de risco ao longo do ciclo de vida dos sistemas de IA (Sadiq et al., 2021; Raji et al., 2020; Ferreira; Grilo; Maia, 2025).

Diante desse contexto, este estudo tem como objetivo desenvolver e aplicar um instrumento integrado de avaliação da maturidade organizacional e da exposição a fontes de risco em inteligência artificial, fundamentado nos requisitos da ISO/IEC 42001 e nas diretrizes da ISO/IEC 23894. Para demonstrar sua aplicabilidade, o instrumento foi aplicado, por meio de um questionário voluntário e anônimo, a organizações públicas e privadas, permitindo verificar sua utilização em um contexto organizacional real.

2 REFERENCIAL TEÓRICO

2.1 INTELIGÊNCIA ARTIFICIAL EM ORGANIZAÇÕES

A utilização de inteligência artificial (IA) tem despertado crescente interesse acadêmico e institucional em organizações de diferentes setores, devido ao potencial dessas tecnologias para automatizar processos, aumentar a eficiência organizacional, ampliar a capacidade de prestação de serviços e gerar valor em diversos contextos de aplicação (Wirtz; Weyerer; Geyer, 2018).

De forma geral, sistemas de IA podem ser definidos como sistemas computacionais capazes de realizar tarefas que normalmente exigiriam inteligência humana, incluindo o reconhecimento de padrões, o aprendizado a partir de dados, a tomada de decisão e o processamento de linguagem natural (Russell; Norvig, 2021). Essas capacidades têm sido aplicadas em contextos organizacionais variados, como a análise de dados financeiros, a detecção de fraudes, a automação de processos, o apoio à formulação de estratégias e a personalização de serviços.

Entretanto, a adoção dessas tecnologias também acarreta desafios relacionados à legitimidade institucional, à proteção de direitos fundamentais e à transparência nos processos decisórios. Como destacam Veale, Van Kleek e Binns (2018), decisões automatizadas por sistemas de IA podem gerar impactos significativos sobre indivíduos e grupos, especialmente quando envolvem processos de alto risco como concessão de crédito, seleção de pessoal, diagnósticos médicos ou aplicação de benefícios públicos.

No contexto brasileiro, iniciativas recentes têm buscado estruturar políticas públicas voltadas ao desenvolvimento e ao uso responsável da inteligência artificial. O Plano Brasileiro de Inteligência Artificial (PBIA) 2024–2028 estabelece diretrizes para pesquisa, infraestrutura computacional, formação de recursos humanos e aplicação da IA em serviços públicos (Brasil, 2024). De forma complementar, o Decreto nº 12.069/2024 instituiu a Estratégia Nacional de Governo Digital para o período de 2024 a 2027, estabelecendo diretrizes para a modernização administrativa, a interoperabilidade de dados e o uso de tecnologias emergentes na administração pública federal (Brasil, 2024).

À medida que a IA é incorporada aos processos organizacionais, torna-se necessária a adoção de mecanismos de governança e gestão que assegurem seu uso responsável, transparente

e alinhado aos objetivos institucionais (ISO, 2023; NIST, 2023).

2.2 GOVERNANÇA E GESTÃO DE INTELIGÊNCIA ARTIFICIAL

A governança de inteligência artificial pode ser definida como o conjunto de estruturas, processos e mecanismos institucionais destinados a orientar, supervisionar e responsabilizar o desenvolvimento e o uso de sistemas de IA (Raji et al., 2020). Esse conceito envolve dimensões estratégicas, como a definição de políticas organizacionais, princípios éticos, estruturas de supervisão e mecanismos de prestação de contas.

A gestão de inteligência artificial refere-se à operacionalização dessas diretrizes no âmbito organizacional, por meio da execução, do monitoramento e do aperfeiçoamento de processos relacionados ao desenvolvimento, à implantação e ao uso de sistemas de IA. A distinção entre governança e gestão é relevante porque muitas iniciativas institucionais concentram-se na formulação de princípios ou diretrizes estratégicas, mas não estabelecem, necessariamente, mecanismos organizacionais capazes de garantir sua implementação efetiva (Raji et al., 2020).

2.3 SISTEMAS DE GESTÃO DE INTELIGÊNCIA ARTIFICIAL E A ISO/IEC 42001

Os sistemas de gestão constituem estruturas organizacionais destinadas a estabelecer políticas, objetivos, processos, responsabilidades e mecanismos de monitoramento para assegurar o alcance de objetivos específicos por meio da melhoria contínua. Aplicados ao contexto da inteligência artificial, esses sistemas buscam institucionalizar práticas que permitam às organizações desenvolver, implantar e utilizar sistemas de IA de forma consistente, responsável e alinhada aos seus objetivos estratégicos (ISO, 2015; ISO, 2023).

A ISO/IEC 42001, publicada em 2023 pela International Organization for Standardization (ISO) em conjunto com a International Electrotechnical Commission (IEC), constitui a primeira norma internacional a especificar requisitos para Sistemas de Gestão de Inteligência Artificial (Artificial Intelligence Management Systems – AIMS). A norma estabelece requisitos que permitem às organizações desenvolver, implementar e manter sistemas de gestão voltados ao uso responsável de tecnologias de IA, seguindo o modelo de alto nível adotado em outras normas de

sistemas de gestão, como a ISO 9001 e a ISO/IEC 27001 (ISO, 2023). Diferentemente de frameworks voltados predominantemente à formulação de princípios ou recomendações, a ISO/IEC 42001 adota a abordagem de sistemas de gestão consolidada pela ISO, estabelecendo requisitos organizacionais auditáveis para institucionalizar a gestão da inteligência artificial.

Entre os principais elementos contemplados pela norma estão a análise do contexto organizacional e das partes interessadas, a definição de políticas e responsabilidades relacionadas à inteligência artificial, a avaliação e o tratamento de riscos associados aos sistemas de IA, a gestão do ciclo de vida das soluções, o monitoramento e a avaliação do desempenho dos sistemas e a promoção da melhoria contínua das práticas organizacionais relacionadas ao uso e à governança da IA (ISO, 2023). Esses requisitos estruturam a gestão da IA segundo o ciclo de melhoria contínua adotado pelas normas ISO de sistemas de gestão, integrando aspectos estratégicos, operacionais e de gestão de riscos ao longo do ciclo de vida dos sistemas de IA.

Contudo, a ISO/IEC 42001 não define níveis de maturidade nem instrumentos específicos para avaliar o grau de desenvolvimento institucional das organizações na implementação dessas práticas. Nesse contexto, modelos de maturidade organizacional podem complementar a norma ao fornecer instrumentos estruturados capazes de avaliar o grau de implementação das práticas previstas, apoiar diagnósticos institucionais e orientar processos de melhoria contínua (Saleh; Abdulsalam, 2024).

2.4 MODELOS DE MATURIDADE ORGANIZACIONAL

Modelos de maturidade organizacional são instrumentos estruturados que permitem avaliar o grau de desenvolvimento de capacidades institucionais em determinado domínio, oferecendo uma progressão de níveis que vai desde práticas ad hoc e pouco estruturadas até estágios de otimização e melhoria contínua. Um dos referenciais mais consolidados nesse campo é o Capability Maturity Model Integration (CMMI), que organiza a maturidade em cinco níveis: inicial, gerenciado, definido, quantitativamente gerenciado e otimizado (CMMI Institute, 2018).

Além de caracterizar o estágio de desenvolvimento das capacidades organizacionais, os modelos de maturidade fornecem uma base estruturada para identificar lacunas, orientar a priorização de ações de melhoria e acompanhar a evolução institucional ao longo do tempo. Por essa razão, são amplamente utilizados como instrumentos de diagnóstico e apoio à melhoria

continua em diferentes domínios da gestão (CMMI Institute, 2018; ISO/IEC, 2003).

No contexto da inteligência artificial, Sadiq et al. (2021), em revisão sistemática da literatura, identificaram que os modelos existentes abordavam dimensões como estratégia, dados, tecnologia, pessoas e governança, mas foram desenvolvidos em um contexto anterior à consolidação de normas específicas para Sistemas de Gestão de IA, o que limita sua integração com referenciais normativos voltados à gestão organizacional da IA. Embora voltado especificamente à IA responsável, o modelo de Ferreira, Grilo e Maia (2025) compartilha com o presente estudo a utilização de um instrumento estruturado de autoavaliação como mecanismo de diagnóstico organizacional e apoio à melhoria contínua

2.5 GESTÃO DE RISCOS EM INTELIGÊNCIA ARTIFICIAL E A ISO/IEC 23894

A ISO/IEC 23894:2023 constitui o referencial normativo internacional específico para a gestão de riscos associados a sistemas de inteligência artificial. Publicada pela ISO em conjunto com a IEC, a norma fornece orientações para a identificação, avaliação, tratamento e monitoramento dos riscos associados aos sistemas de IA, integrando a gestão de riscos aos processos organizacionais, sendo compatível com a estrutura geral da ISO 31000 e com os requisitos da ISO/IEC 42001 (ISO, 2023b).

O Anexo B da norma identifica sete fontes de risco características de sistemas de IA: (1) complexidade do ambiente de operação; (2) limitações de transparência e explicabilidade; (3) grau de automação e supervisão humana insuficiente; (4) riscos associados a dados e ao aprendizado de máquina; (5) dependências de infraestrutura e hardware; (6) lacunas no ciclo de vida do sistema; e (7) inadequação no nível de prontidão tecnológica (ISO, 2023b). Diferentemente de riscos tecnológicos convencionais, essas fontes decorrem de características intrínsecas da IA, como o aprendizado baseado em dados, o comportamento adaptativo, a baixa explicabilidade e o elevado grau de autonomia.

A articulação entre a ISO/IEC 23894 e a ISO/IEC 42001 é explicitamente prevista pelo próprio arcabouço normativo: a 42001 determina que as organizações realizem a avaliação e o tratamento de riscos de IA como parte do seu sistema de gestão (seção 6.1), enquanto a 23894 fornece diretrizes e categorias de risco que apoiam a operacionalização desse requisito. Essa complementaridade justifica, do ponto de vista conceitual, a integração das duas normas em um

único instrumento de avaliação, que permita verificar simultaneamente o grau de implementação dos controles (maturidade) e o nível de exposição declarada às fontes de risco específicas da IA (ISO, 2023; ISO, 2023b). Essa integração permite associar a capacidade organizacional de gerenciar sistemas de IA ao contexto de riscos enfrentado pela organização, fornecendo uma avaliação mais abrangente do que a obtida pela utilização isolada de cada um dos referenciais.

3 METODOLOGIA

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa caracteriza-se como aplicada, com abordagem quantitativa e caráter exploratório. O foco central não é a análise estatística de uma amostra representativa de organizações, mas sim o desenvolvimento, a descrição e a validação preliminar de um instrumento integrado de avaliação da maturidade organizacional e da exposição a fontes de risco para a gestão de inteligência artificial. O instrumento foi aplicado em um conjunto inicial de nove organizações com o propósito de verificar sua aplicabilidade operacional, a coerência interna do modelo e a capacidade de gerar diagnósticos individuais inteligíveis, e não para produzir inferências estatísticas sobre o universo de organizações (Wazlawick, 2020).

O desenvolvimento do instrumento seguiu uma abordagem dedutiva, partindo dos referenciais normativos da ISO/IEC 42001 e da ISO/IEC 23894 para a construção dos blocos de avaliação e do modelo de mapeamento entre controles e fontes de risco. A aplicação em organizações de perfis variados, públicas e privadas, de diferentes portes e estágios de adoção de IA, serviu como teste de aderência prática do instrumento, permitindo verificar se os itens eram compreensíveis, se a escala era operacionalizável pelos respondentes e se o cruzamento entre os dois blocos gerava resultados analiticamente coerentes.

3.2 CONSTRUÇÃO DO MODELO DE MATURIDADE

O instrumento integrado foi desenvolvido a partir da articulação de dois referenciais normativos complementares: os requisitos da ISO/IEC 42001 (ISO, 2023), que orientam a construção do modelo de maturidade de controles, e as fontes de risco do Anexo B da ISO/IEC

23894 (ISO, 2023b), que fundamentam a avaliação do grau de exposição declarada da organização a cada fonte.

Importa precisar o sentido em que o termo "maturidade" é empregado neste estudo. Em consonância com a tradição de modelos de maturidade aplicados à IA (Sadiq et al., 2021; Ferreira; Grilo; Maia, 2025), a maturidade é aqui operacionalizada como o grau de implementação autodeclarado dos controles do Anexo A da ISO/IEC 42001, mensurado pela escala NPLF. Trata-se de uma medida de percepção organizacional, sem verificação por evidências externas, auditoria independente ou validação cumulativa entre níveis, tal como ocorre em modelos de maturidade auditáveis como o CMMI (CMMI Institute, 2018). Essa delimitação é deliberada: o instrumento proposto destina-se ao autodiagnóstico organizacional, e não à certificação ou à auditoria de conformidade.

O instrumento é estruturado em três blocos. O Bloco 1 destina-se à identificação e caracterização da organização respondente. O Bloco 2 (Avaliação de Riscos) é composto por sete itens correspondentes às fontes de risco do Anexo B da ISO/IEC 23894, mensurados em escala de 1 a 5, de acordo com o grau de exposição declarado. O Bloco 3 (Avaliação de Controles) é composto por trinta e oito itens derivados dos controles do Anexo A da ISO/IEC 42001, organizados em seis dimensões de controle e respondidos em escala NPLF. Para assegurar a rastreabilidade entre fontes de risco e controles, cada controle foi associado às fontes de risco que contribui para mitigar, conforme o Quadro 1.

Quadro 1. Fontes de risco do Anexo B da ISO/IEC 23894 e controles correspondentes do Anexo A da ISO/IEC 42001.

Fonte de Risco (ISO/IEC 23894, Anexo B)	Controles ISO/IEC 42001 (Anexo A) Relacionados
R1. Complexidade do ambiente	A.4.2, A.5.2, A.6.2.2, A.6.2.3, A.6.2.4, A.6.2.6
R2. Transparência e explicabilidade	A.2.2, A.2.3, A.5.3, A.5.4, A.5.5, A.6.2.7, A.6.2.8, A.8.2, A.8.3, A.8.4, A.8.5
R3. Nível de automação	A.3.2, A.3.3, A.4.6, A.5.2, A.9.2, A.9.3, A.9.4, A.10.2, A.10.4
R4. Dados e aprendizado de máquina	A.2.3, A.4.3, A.6.2.4, A.6.2.6, A.7.2, A.7.3, A.7.4, A.7.5, A.7.6
R5. Infraestrutura e hardware	A.4.5, A.6.2.5, A.6.2.6, A.10.3
R6. Ciclo de vida do sistema	A.2.2, A.2.4, A.6.1.2, A.6.1.3, A.6.2.2, A.6.2.3, A.6.2.4, A.6.2.5, A.6.2.6, A.6.2.8, A.10.2
R7. Prontidão tecnológica	A.2.4, A.4.4, A.5.2, A.6.1.2, A.6.2.2, A.10.3

Fonte: Elaborado pelos autores.

A correspondência entre cada fonte de risco e os controles do Anexo A foi construída a partir da análise dos próprios referenciais normativos, assegurando a validade de conteúdo do modelo: cada controle do Anexo A encontra justificativa normativa para sua associação à(s) fonte(s) de risco correspondente(s) no Anexo B (Saleh; Abdulsalam, 2024). Essa rastreabilidade é o elemento central da proposta, pois permite ao instrumento não apenas medir isoladamente o nível de maturidade dos controles ou a percepção de exposição a fontes de risco, mas cruzar essas duas dimensões de forma estruturada e normativamente fundamentada.

3.3 INSTRUMENTO DE COLETA DE DADOS E ESCALAS DE MENSURAÇÃO

O Bloco 2 utiliza uma escala Likert de 5 pontos para a avaliação de exposição às fontes de risco: (1) Muito baixa; (2) Baixa; (3) Moderada; (4) Alta; e (5) Muito alta. Essa escala permite ao respondente posicionar a organização em relação a cada fonte de risco de forma simples e direta, sem exigir conhecimento técnico aprofundado das normas, o que é uma característica importante para a acessibilidade do instrumento a diferentes perfis de respondentes.

O Bloco 3 utiliza a escala NPLF, originalmente definida pela ISO/IEC 15504-2:2003 e adotada pelo COBIT 5 (ISACA, 2012): (N) Não atingido — de 0% a 15% de alcance do controle; (P) Parcialmente atingido — de 15% a 50%; (L) Largamente atingido — de 50% a 85%; e (F) Completamente atingido — de 85% a 100%. A adoção de uma escala normativamente consolidada, em vez de uma escala ad hoc, confere ao instrumento maior credibilidade metodológica e facilita eventual comparação com outros estudos que utilizem o mesmo referencial (ISO/IEC, 2003).

O cruzamento entre os dois blocos é realizado por meio de uma análise comparativa entre o nível de exposição declarado a cada fonte de risco (Bloco 2) e o grau de implementação dos controles normativamente associados a essa fonte (Bloco 3). Para cada par fonte de risco / organização, o instrumento posiciona simultaneamente a percepção de exposição e a maturidade dos controles correspondentes, permitindo identificar situações de alinhamento — quando a maturidade dos controles é proporcional à exposição percebida — e situações de lacuna de controle — quando a exposição declarada é elevada e a maturidade dos controles associados é baixa. Essa análise comparativa não constitui uma medida formal de risco residual no sentido da ISO 31000, uma vez que o instrumento não realiza avaliação formal de risco, mas oferece uma

visão diagnóstica estruturada da relação entre o que a organização percebe como fonte de exposição e o quanto ela declarou ter avançado na implementação dos controles normativamente relacionados.

O instrumento foi operacionalizado por meio de formulário eletrônico na plataforma Google Forms, respondido preferencialmente por profissional responsável pela governança de TI ou de IA da organização, ou por equipe multidisciplinar com conhecimento das práticas avaliadas.

3.4 PROCEDIMENTOS DE APLICAÇÃO E LIMITAÇÕES DO ESTUDO PILOTO

O instrumento foi disponibilizado a diversos grupos de profissionais ligados à governança e à gestão de tecnologia da informação, com adesão voluntária das organizações respondentes. As nove organizações que responderam ao instrumento não foram selecionadas por critérios prévios de controle amostral, mas aderiram espontaneamente ao convite de participação. Esse conjunto, embora reduzido e não probabilístico, reuniu perfis variados em termos de natureza institucional (cinco públicas e quatro privadas), porte (quatro grandes, duas médias e três pequenas) e estágio de adoção de IA (cinco em produção, duas em desenvolvimento e duas com planos formais), o que contribuiu para testar o instrumento em contextos organizacionais distintos.

É importante delimitar o escopo analítico desta aplicação. Dado o número reduzido de respondentes e o método de recrutamento por adesão voluntária, os dados obtidos não permitem inferências estatísticas sobre a maturidade ou a exposição a riscos de IA em setores ou tipos de organização. A aplicação piloto teve finalidade instrumental: verificar se o instrumento é compreensível e operacionalizável por respondentes de diferentes perfis, se as escalas funcionam conforme esperado e se o cruzamento entre os blocos gera saídas analiticamente coerentes. A análise dos resultados, apresentada na seção seguinte, é orientada pela perspectiva de validação do instrumento como método, e não pela busca de conclusões sobre o universo de organizações.

4 RESULTADOS E DISCUSSÕES

4.1 APLICABILIDADE OPERACIONAL DO INSTRUMENTO

O instrumento foi respondido integralmente pelas nove organizações participantes, sem ocorrência de recusas de itens, abandono do formulário ou necessidade de esclarecimentos adicionais por parte dos respondentes. Todos os quarenta e cinco itens (sete do Bloco 2 e trinta e oito do Bloco 3) foram respondidos em todas as submissões, o que indica que o instrumento é operacionalizável na extensão em que foi proposto, sem gerar sobrecarga excessiva ou ambiguidade que inviabilizasse o preenchimento.

As escalas utilizadas (Likert de 1 a 5 para o Bloco 2 e NPLF para o Bloco 3) foram empregadas em toda a sua extensão pelos respondentes: todos os cinco pontos da escala de exposição foram utilizados, assim como todos os quatro níveis da escala NPLF, evidenciando que as âncoras das escalas foram compreendidas e que os respondentes não tenderam à centralidade ou à extremidade de forma sistemática. Isso sugere que as escalas discriminam adequadamente diferentes percepções e situações organizacionais, o que é uma propriedade desejável em instrumentos de autoavaliação (Sadiq et al., 2021).

O perfil das organizações participantes, apresentado no Quadro 2, ilustra a diversidade de contextos em que o instrumento foi testado.

Quadro 2. Perfil das organizações participantes na aplicação-piloto.

Organização	Natureza	Setor de Atividade	Porte	Estágio de Adoção de IA
Org1	Privada	Tecnologia / Telecomunicações	Pequena	Planos formais para adoção
Org2	Pública	Administração pública / Governo	Grande	Em produção
Org3	Privada	Tecnologia / Telecomunicações	Pequena	Em produção
Org4	Pública	Administração pública / Governo	Grande	Desenvolvimento / piloto
Org5	Privada	Tecnologia / Telecomunicações	Pequena	Desenvolvimento / piloto
Org6	Privada	Saúde	Grande	Em produção
Org7	Pública	Administração pública / Governo	Grande	Em produção
Org8	Pública	Administração pública / Governo	Média	Em produção
Org9	Pública	Educação	Média	Planos formais para adoção

Fonte: Elaborado pelos autores.

A aplicação do instrumento em organizações de portes, naturezas, setores e estágios de adoção de IA distintos permitiu verificar que ele funciona adequadamente em contextos heterogêneos, sem depender de características específicas do respondente ou da organização para gerar resultados. Essa versatilidade é relevante para a proposta do instrumento como ferramenta de uso amplo.

4.2 COERÊNCIA INTERNA DO MODELO DE MAPEAMENTO

Uma das questões centrais a verificar na aplicação piloto era se o cruzamento entre os Blocos 2 e 3, gerava saídas analiticamente coerentes com a lógica do mapeamento normativo proposto no Quadro 1. Em outras palavras: organizações que declararam maior exposição a determinada fonte de risco tenderam a apresentar menor maturidade nos controles normativamente associados a essa fonte?

Os resultados da aplicação piloto apontam para uma coerência interna satisfatória do modelo. A Tabela 1 apresenta, para cada organização participante, o nível de exposição declarada à fonte de risco R1 (Complexidade do ambiente) e o nível de maturidade correspondente dos controles a ela associados, a título ilustrativo do funcionamento do cruzamento proposto pelo instrumento.

Tabela 1. Ilustração do cruzamento entre exposição declarada e maturidade dos controles, por organização, para a fonte R1, Complexidade do ambiente.

Organização	Natureza	Exposição Declarada (1–5)	Maturidade dos Controles (NPLF)
Org2	Pública	3 – Moderada	L – Largamente atingido
Org4	Pública	4 – Alta	P – Parcialmente atingido
Org6	Privada	3 – Moderada	L – Largamente atingido
Org7	Pública	3 – Moderada	P – Parcialmente atingido
Org8	Pública	4 – Alta	N – Não atingido
Org9	Pública	5 – Muito alta	N – Não atingido
Org1	Privada	2 – Baixa	N – Não atingido
Org3	Privada	1 – Muito baixa	N – Não atingido
Org5	Privada	2 – Baixa	N – Não atingido

Fonte: Elaborado pelos autores.

Os dados ilustrados na Tabela 1 evidenciam o padrão esperado pela lógica do instrumento: organizações que declararam maior exposição à fonte R1 e que, simultaneamente, apresentaram níveis mais baixos de maturidade dos controles correspondentes, configuram as situações de maior lacuna de controle identificadas na aplicação piloto. Org9, com exposição máxima (5) e maturidade no nível N (não atingido), representa o caso de maior desalinhamento entre a exposição percebida e os controles implementados. Em contraste, Org2 e Org6, com exposição moderada (3) e maturidade no nível L (largamente atingido), ilustram situações de alinhamento, em que o grau de implementação dos controles é proporcional à exposição declarada. Esse padrão de coerência entre exposição e maturidade se reproduziu de forma análoga nas demais fontes de risco avaliadas, sustentando a avaliação de coerência interna do modelo.

De forma complementar, verificou-se que o instrumento é capaz de identificar situações em que uma organização apresenta alta exposição declarada a determinada fonte de risco combinada com baixa maturidade dos controles correspondentes, configuração que o modelo denomina lacuna de controle e que representa a informação diagnóstica mais relevante para a tomada de decisão organizacional. Essa capacidade diagnóstica é o atributo central que justifica a proposta do instrumento como ferramenta de apoio à gestão de IA.

4.3 DISCUSSÃO SOBRE A ADEQUAÇÃO DO INSTRUMENTO COMO MÉTODO

A aplicação piloto evidenciou que a estrutura em três blocos, com dois blocos analíticos articulados por um mapeamento normativo rastreável, é adequada ao objetivo de autodiagnóstico organizacional. A separação entre a avaliação de exposição a riscos (Bloco 2) e a avaliação de maturidade de controles (Bloco 3) permite que a organização visualize não apenas onde está mais vulnerável, mas também onde os controles implementados são insuficientes em relação à exposição percebida, informação que uma avaliação de maturidade isolada não poderia fornecer.

Essa abordagem guarda convergência com a proposta de Ferreira, Grilo e Maia (2025), que desenvolveram um modelo de maturidade para a adoção responsável de IA, estruturado como instrumento de autoavaliação, capaz de gerar diagnósticos e orientar processos de melhoria. A distinção central entre os dois trabalhos reside no referencial normativo adotado: enquanto Ferreira, Grilo e Maia (2025) constroem seu modelo a partir de dimensões conceituais mais

amplas, o instrumento aqui proposto ancora-se explicitamente nos Anexos A e B de duas normas ISO complementares, o que lhe confere rastreabilidade normativa direta e facilita sua adoção por organizações que já operam em ambientes regulados por referenciais ISO. De forma análoga, Sadiq et al. (2021), em revisão sistemática da literatura sobre modelos de maturidade em IA, identificaram como lacuna recorrente a ausência de referenciais normativos consolidados como base para a construção desses modelos, lacuna que o presente instrumento busca endereçar ao fundamentar-se explicitamente na ISO/IEC 42001 e na ISO/IEC 23894. Por sua vez, Saleh e Abdulsalam (2024), ao analisarem a operacionalização da ISO/IEC 42001, apontam a dificuldade de traduzir seus requisitos em critérios verificáveis de conformidade, o que reforça a pertinência de instrumentos intermediários, como o aqui proposto, que operam na camada de autoavaliação sem reivindicar o rigor de uma auditoria formal.

A escolha pela escala NPLF, em vez de escalas numéricas ou de frequência, demonstrou-se adequada para o contexto de autoavaliação da implementação de controles normativos. As âncoras descritivas da escala, ancoradas em faixas percentuais de alcance definidas pela ISO/IEC 15504-2:2003, oferecem ao respondente critérios objetivos para o posicionamento, reduzindo a subjetividade inerente a escalas de concordância genéricas. Essa propriedade é especialmente relevante quando o instrumento é respondido por profissionais de diferentes formações e níveis de familiaridade com as normas de referência.

O instrumento revelou, ainda, que a granularidade do mapeamento, com trinta e oito controles distribuídos em seis dimensões e associados a sete fontes de risco, é suficiente para gerar diagnósticos diferenciados entre organizações de perfis distintos, sem ser excessivamente extenso a ponto de inviabilizar o preenchimento. Essa adequação entre abrangência e praticidade é uma das contribuições metodológicas do instrumento proposto.

Do ponto de vista das implicações práticas, o instrumento destina-se a gestores de TI, responsáveis por governança de IA e equipes de conformidade. É útil tanto antes da implementação de sistemas de IA, como ferramenta de diagnóstico inicial, quanto após implementações já em curso, como mecanismo de identificação de lacunas emergentes. A comparação entre a exposição declarada e a maturidade dos controles oferece uma visão priorizada das áreas que demandam atenção, permitindo que organizações com recursos limitados concentrem seus esforços nas lacunas de maior relevância.

Por outro lado, a aplicação piloto também evidenciou limitações inerentes ao método de autoavaliação: os resultados refletem a percepção do respondente sobre o estado dos controles e dos riscos, e não uma verificação objetiva de conformidade. Organizações que não dispõem de profissionais familiarizados com as normas de referência podem ter dificuldades em avaliar adequadamente o nível de implementação de controles técnicos específicos. Essa limitação aponta para a necessidade de acompanhar futuras aplicações do instrumento com orientações de preenchimento mais detalhadas, ou de combiná-lo com métodos de verificação documental para contextos em que maior rigor seja necessário.

Para a agenda de pesquisa futura, os resultados deste estudo apontam três direções prioritárias. A primeira é a validação do instrumento com amostras maiores e de recrutamento controlado, que permitam verificar se as propriedades de aplicabilidade e de coerência interna se mantêm em escala e se emergem padrões setoriais ou institucionais com significância estatística. A segunda é a combinação do instrumento de autoavaliação com métodos de verificação documental ou auditoria, de modo a avaliar a correspondência entre a percepção declarada de maturidade e a implementação efetiva dos controles. A terceira é a adaptação do instrumento a setores ou contextos específicos, como a administração pública federal, o setor de saúde ou o setor financeiro, nos quais os riscos e os requisitos regulatórios de IA apresentam características particulares que podem demandar ajustes no mapeamento entre fontes de risco e controles.

5 CONCLUSÃO

Este estudo partiu de um problema identificado na interface entre a normatização internacional de inteligência artificial e a prática organizacional: apesar da publicação de referenciais normativos estruturados, em especial a ISO/IEC 42001, que estabelece requisitos para a gestão de IA, e a ISO/IEC 23894, que fornece orientação para a gestão de riscos de IA, não se encontra na literatura um instrumento integrado capaz de articular sistematicamente a avaliação da maturidade dos controles de gestão com a exposição declarada às fontes de risco identificadas nessas normas. Esse foi o problema que motivou o desenvolvimento do instrumento proposto.

O instrumento construído responde a essa lacuna ao articular, em uma única ferramenta de autoavaliação, os trinta e oito controles do Anexo A da ISO/IEC 42001, organizados em seis

dimensões de controle, com as sete fontes de risco do Anexo B da ISO/IEC 23894, por meio de um mapeamento normativo rastreável. Esse mapeamento constitui a principal contribuição metodológica do estudo: ao associar explicitamente cada controle às fontes de risco que ele contribui para mitigar, o instrumento torna visível a relação entre o que a organização declara implementar e o que ela declara enfrentar como fontes de risco, permitindo identificar, de forma estruturada e normativamente fundamentada, lacunas de controle: situações em que a exposição percebida é alta e a maturidade dos controles correspondentes é baixa.

A aplicação piloto do instrumento em nove organizações de perfis distintos, públicas e privadas, de diferentes portes, setores e estágios de adoção de IA, permitiu verificar três propriedades essenciais para sua utilidade como ferramenta diagnóstica. Primeiro, sua aplicabilidade operacional: o instrumento foi respondido integralmente em todos os casos, sem ambiguidades ou recusas de itens, e as escalas foram utilizadas em toda a sua extensão, indicando que são compreensíveis e discriminativas para respondentes de diferentes perfis. Segundo sua coerência interna: o cruzamento entre os dois blocos analíticos gerou saídas consistentes com a lógica do mapeamento normativo, com organizações de maior exposição declarada tendendo a apresentar menor maturidade nos controles associados, o que valida a racionalidade do modelo. Terceiro, sua versatilidade: o instrumento funcionou adequadamente em contextos heterogêneos, sem depender de características específicas da organização ou do respondente para gerar resultados analiticamente inteligíveis.

Do ponto de vista das limitações, é necessário reconhecer que o instrumento produz medidas de percepção organizacional autodeclarada, e não verificações objetivas de conformidade. A ausência de auditoria externa, de validação cumulativa entre níveis e de checagem documental implica que os resultados devem ser interpretados como ponto de partida diagnóstico, e não como atestado de conformidade com as normas de referência. Adicionalmente, a aplicação piloto envolveu um número reduzido de organizações, selecionadas por adesão voluntária, o que inviabiliza qualquer inferência estatística sobre o universo de organizações. Essas limitações são inerentes ao escopo e ao objetivo do estudo, que é desenvolver e verificar um instrumento e não caracterizar empiricamente uma população, e devem orientar a interpretação dos resultados e o desenho de estudos futuros.

Em síntese, o instrumento proposto oferece uma contribuição metodológica concreta ao campo da gestão de IA: um método estruturado, normativamente fundamentado e praticamente

acessível para o autodiagnóstico organizacional da maturidade dos controles e da exposição a fontes de risco em inteligência artificial, aplicável a organizações de qualquer natureza, porte e setor. Em um contexto em que a adoção de sistemas de IA avança mais rapidamente do que as estruturas de governança correspondentes, dispor de instrumentos capazes de orientar diagnósticos e priorizar ações de melhoria representa uma contribuição relevante tanto para a comunidade acadêmica quanto para os gestores e tomadores de decisão responsáveis pelo uso responsável da inteligência artificial nas organizações.



REFERÊNCIAS

BRASIL. Ministério da Ciência, Tecnologia e Inovação. **Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028**. Brasília: MCTI, 2024. Disponível em: <https://www.gov.br/lncc/pt-br/assuntos/noticias/ultimas-noticias-1/plano-brasileiro-de-inteligencia-artificial-pbia-2024-2028>. Acesso em: 9 mar. 2026.

BRASIL. Presidência da República. **Decreto nº 12.069, de 21 de fevereiro de 2024**. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/decreto/D12069.htm. Acesso em: 9 mar. 2026.

CMMI INSTITUTE. **CMMI for Development**, Version 2.0. Pittsburgh: CMMI Institute, 2018. Disponível em: <https://cmmiinstitute.com>. Acesso em: 22 mar. 2026.

EUROPEAN UNION. **Artificial Intelligence Act: Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence**. Brussels: European Union, 2024. Disponível em: <https://artificial-intelligence-act.eu/>. Acesso em: 9 mar. 2026.

FERREIRA, R. M. F. D.; GRILO, A.; MAIA, M. Piloting a maturity model for responsible artificial intelligence: a Portuguese case study. **Journal of Responsible Technology**, Amsterdam, v. 22, 100117, jan. 2025.

ISACA. **COBIT 5: a business framework for the governance and management of enterprise IT**. Rolling Meadows: ISACA, 2012.

ISO – INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 15504-2:2003**. Information technology. Process assessment. Part 2: performing an assessment. Geneva: ISO, 2003.

ISO – INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 42001:2023**. Information technology. Artificial intelligence. Management system. Geneva: ISO, 2023a.

ISO – INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 23894:2023**. Information technology. Artificial intelligence. Guidance on risk management. Geneva: ISO, 2023b.

NIST – NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**. Gaithersburg: NIST, 2023. Disponível em: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>. Acesso em: 9 mar. 2026.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD Principles on Artificial Intelligence**. Paris: OECD, 2019. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 9 mar. 2026.

RAJI, I. D. et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: **CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT)**, 2020, New York. Proceedings. New York: ACM, 2020.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: a modern approach**. 4. ed. Upper Saddle River: Pearson, 2021.

SADIQ, R. B. et al. Artificial intelligence maturity model: a systematic literature review. **PeerJ Computer Science**, San Diego, v. 7, e661, ago. 2021.

SALEH, K.; ABDULSALAM, H. Operationalizing ISO/IEC 42001: requirements and conformance evidence for AI management systems. In: **INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE FOR SUSTAINABLE INNOVATION (AI-SI)**, 2025, [s.l.]. Proceedings. [S.l.]: IEEE, 2025.

VEALE, M.; VAN KLEEK, M.; BINNS, R. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In: **CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS**, 2018, New York. Proceedings. New York: ACM, 2018.

WAZLAWICK, R. S. **Metodologia de pesquisa para ciência da computação**. 3. ed. Rio de Janeiro: Elsevier, 2020.

WIRTZ, B. W.; WEYERER, J. C.; GEYER, C. Artificial intelligence and the public sector: applications and challenges. **International Journal of Public Administration**, Abingdon, v. 42, n. 7, p. 596-615, jun. 2018.