DOI: 10.55905/rdelosv19.n84-001
ISSN: 1988-5245

Originals received: 7/3/2026
Acceptance for publication: 7/23/2026

Contribuição das medidas do PPSI 2.0 para o tratamento de riscos de Inteligência Artificial na Administração Pública Federal

Contribution of PPSI 2.0 measures to the treatment of Artificial Intelligence risks in the brazilian Federal Public Administration

Contribución de las medidas del PPSI 2.0 al tratamiento de los riesgos de Inteligencia Artificial en la Administración Pública Federal brasileña

Ana Paula Morais do Vale
Mestranda em Segurança Cibernética
Instituição: Universidade de Brasília (UNB)
Endereço: Brasília - Distrito Federal, Brasil
E-mail: apmv.prf@gmail.com

João Souza Neto
Pós-Doutor em Segurança Cibernética
Instituição: Universidade de Brasília (UNB)
Endereço: Brasília - Distrito Federal, Brasil
E-mail: szneto@gmail.com

RESUMO
A incorporação da inteligência artificial (IA) à Administração Pública Federal amplia riscos de segurança, privacidade, desempenho, direitos e governança, exigindo avaliar a adequação dos instrumentos institucionais existentes. Este estudo avaliou, em termos documentais e causais, a contribuição das medidas do PPSI 2.0 para o tratamento de riscos de IA relevantes para a APF. A pesquisa, aplicada e documental, normalizou 158 registros extraídos de referenciais técnicos e normativos em 87 cenários de risco. Em seguida, as 210 medidas do PPSI foram confrontadas com esses cenários em uma estrutura muitos-para-muitos, classificando-se as relações como diretas, parciais, indiretas ou inexistentes. Foram identificadas 921 relações materiais, com predominância das parciais (607). Apenas 35 riscos apresentaram alguma medida diretamente aplicável, enquanto 52 não obtiveram cobertura direta. As principais lacunas concentraram-se na validação e no monitoramento de modelos, na segurança adversarial, nos direitos e na supervisão humana e nas competências. A classificação das contribuições como diretas, parciais ou indiretas evita dois erros opostos: desconsiderar a utilidade das medidas já institucionalizadas por não terem sido concebidas especificamente para a IA e presumir que práticas gerais de privacidade e segurança sejam suficientes para tratar riscos próprios ou intensificados por essa tecnologia. Conclui-se que o PPSI constitui uma base institucional relevante para o tratamento de riscos de IA, mas requer complementações específicas.

Palavras-chave: Inteligência Artificial, gestão de riscos, PPSI 2.0, administração pública, governança digital.

ABSTRACT
The incorporation of artificial intelligence (AI) into the Brazilian Federal Public Administration increases risks related to security, privacy, performance, rights, and governance, requiring an assessment of the adequacy of existing institutional instruments. This study assessed, from documentary and causal perspectives, the contribution of PPSI 2.0 measures to the treatment of AI risks relevant to the Brazilian Federal Public Administration. This applied documentary research normalized 158 records extracted from technical and normative reference documents into 87 risk scenarios. Subsequently, the 210 PPSI measures were mapped against these scenarios using a many-to-many structure, and the relationships were classified as direct, partial, indirect, or nonexistent. A total of 921 substantive relationships were identified, predominantly partial (607). Only 35 risks had at least one directly applicable measure, whereas 52 had no direct coverage. The main gaps concerned model validation and monitoring, adversarial security, rights and human oversight, and competencies. Classifying contributions as direct, partial, or indirect avoids two opposing errors: disregarding the usefulness of measures already institutionalized because they were not specifically designed for AI, and assuming that general privacy and security practices are sufficient to address risks specific to or intensified by this technology. It is concluded that PPSI provides a relevant institutional foundation for treating AI risks but requires AI-specific enhancements.

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

RESUMEN
La incorporación de la inteligencia artificial (IA) a la Administración Pública Federal brasileña amplía los riesgos relacionados con la seguridad, la privacidad, el desempeño, los derechos y la gobernanza, lo que exige evaluar la adecuación de los instrumentos institucionales existentes. Este estudio evaluó, en términos documentales y causales, la contribución de las medidas del PPSI 2.0 al tratamiento de los riesgos de IA relevantes para la Administración Pública Federal brasileña. La investigación, de carácter aplicado y documental, normalizó 158 registros extraídos de referentes técnicos y normativos en 87 escenarios de riesgo. Posteriormente, las 210 medidas del PPSI se contrastaron con estos escenarios mediante una estructura de relaciones muchos a muchos, clasificadas como directas, parciales, indirectas o inexistentes. Se identificaron 921 relaciones sustantivas, con predominio de las parciales (607). Solo 35 riesgos presentaron al menos una medida directamente aplicable, mientras que 52 carecieron de cobertura directa. Las principales brechas se concentraron en la validación y el monitoreo de modelos, la seguridad adversarial, los derechos y la supervisión humana, y las competencias. La clasificación de las contribuciones como directas, parciales o indirectas evita dos errores opuestos: desconsiderar la utilidad de las medidas ya institucionalizadas por no haber sido concebidas específicamente para la IA y presumir que las prácticas generales de privacidad y seguridad son suficientes para tratar riesgos propios de esta tecnología o intensificados por ella. Se concluye que el PPSI constituye una base institucional relevante para el tratamiento de los riesgos de IA, pero requiere complementaciones específicas.

Palabras clave: Inteligencia Artificial, gestión de riesgos, PPSI 2.0, administración pública, gobernanza digital.

1 INTRODUÇÃO

A transformação digital do setor público amplia o uso de dados e de tecnologias emergentes na automação de processos, no apoio à decisão e na prestação de serviços. Nesse contexto, a inteligência artificial (IA) pode fortalecer a formulação de políticas públicas, a gestão interna e a interação entre Estado e sociedade (Wirtz; Weyerer; Geyer, 2019). Sua adoção pela Administração Pública, contudo, ainda é recente e heterogênea. A literatura permanece predominantemente exploratória, conceitual ou baseada em casos específicos, sem evidências longitudinais suficientes para avaliar de forma abrangente seus efeitos institucionais (Sun; Medaglia, 2019; Zuiderwijk; Chien; Salem, 2021). Na Administração Pública Federal brasileira (APF), iniciativas voltadas a ferramentas, capacitações, orientações e *frameworks* indicam que a adoção da IA avança paralelamente à construção de capacidades para seu uso responsável (Brasil, 2026a).

Embora favoreça eficiência, inovação e geração de valor público, a IA introduz riscos decorrentes dos dados, modelos, infraestrutura, interações humano–máquina e contextos organizacionais. Suas consequências alcançam segurança, privacidade, desempenho, transparência, discriminação, responsabilização e proteção de direitos. Esses desafios envolvem dimensões tecnológicas, organizacionais, gerenciais, jurídicas, políticas, éticas, sociais e relacionadas aos dados (Wirtz; Weyerer; Geyer, 2019; Sun; Medaglia, 2019). Por isso, sua governança deve ser orientada a riscos e proporcional às características, finalidades e possíveis consequências de cada aplicação, evitando tanto a adoção desprotegida quanto requisitos dissociados do contexto.

Na APF, o Programa de Privacidade e Segurança da Informação (PPSI 2.0), instituído pela Portaria SGD/MGI nº 9.511, de 28 de outubro de 2025, orienta os órgãos e as entidades integrantes do Sistema de Administração dos Recursos de Tecnologia da Informação (SISP). Seu Framework de Privacidade e Segurança da Informação reúne 210 medidas sobre governança, gestão de riscos e ativos, identidade e acesso, proteção de dados pessoais, segurança de redes, desenvolvimento seguro, fornecedores, continuidade e resposta a incidentes (Brasil, 2026b). Fundamentado em normas e modelos reconhecidos — séries ISO/IEC 27001, 27002 e 27701, NIST Cybersecurity Framework, NIST Privacy Framework, CIS Controls e Secure Controls

Framework —, o PPSI pode constituir uma linha de base institucional para tratar parte dos riscos associados à IA.

O PPSI 2.0, entretanto, não foi concebido como perfil específico de governança de riscos de IA. A existência de medidas gerais sobre dados, acessos, vulnerabilidades, fornecedores e incidentes não demonstra que fenômenos próprios ou intensificados pela IA estejam integralmente contemplados. Deriva de modelos, alucinações, ataques adversariais, explicabilidade, supervisão humana, validação contínua e monitoramento de desempenho podem exigir adaptações ou requisitos adicionais. Permanece, assim, a lacuna de verificar sistematicamente em que medida as 210 medidas contribuem com o tratamento de riscos de IA relevantes para a APF e quais domínios recebem apenas cobertura parcial.

Diante disso, pergunta-se: em que medida as medidas do PPSI 2.0 contribuem para o tratamento dos riscos de inteligência artificial relevantes para a Administração Pública Federal e em quais domínios são necessárias complementações específicas? O estudo objetiva avaliar, com base em critérios de compatibilidade documental e causal, a contribuição das 210 medidas do PPSI 2.0 para o tratamento de 87 cenários normalizados de risco de IA, bem como identificar famílias temáticas que demandam fortalecimento. Neste estudo, a contribuição corresponde à compatibilidade entre o objeto, o mecanismo de atuação ou a finalidade da medida e os componentes do cenário de risco. A classificação resultante não constitui evidência de implementação, eficácia operacional, redução do risco residual ou maturidade institucional.

A contribuição do artigo é uma ponte auditável entre uma taxonomia homogênea de riscos de IA e medidas concretas adotadas pela APF. Ao distinguir relações diretas, parciais, indiretas e inexistentes, a análise evita tanto desconsiderar controles já institucionalizados por não terem sido concebidos para IA quanto presumir que práticas gerais de privacidade e segurança cobrem integralmente fenômenos próprios ou intensificados por essa tecnologia. O estudo limita-se à cobertura documental potencial do PPSI 2.0, sem examinar sua implementação em organizações específicas, estimar probabilidade e impacto ou definir prioridades de tratamento. Além desta introdução, o artigo apresenta o referencial teórico, o método, os resultados, a discussão e as conclusões.

2 REFERENCIAL TEÓRICO

2.1 CONVERGÊNCIA ENTRE NIST AI RMF, ISO/IEC 23894 E ISO/IEC 42001 SOBRE A GESTÃO CONTÍNUA DE RISCOS DE IA

A gestão de riscos de inteligência artificial distingue-se da gestão convencional de riscos tecnológicos por abranger, de forma integrada, componentes técnicos, processos organizacionais, pessoas afetadas e contexto sociotécnico. Esses riscos podem surgir ou se modificar ao longo do uso em razão de mudanças nos dados, no ambiente operacional, nas integrações, no comportamento dos usuários ou no desempenho do modelo. Assim, avaliações anteriores à implantação não bastam para assegurar a confiabilidade durante todo o ciclo de vida, que depende da articulação contínua entre dados, responsabilidades, supervisão e prestação de contas (Janssen *et al*., 2020; Mantymaki *et al*., 2022).

O NIST AI RMF 1.0 estrutura essa gestão em quatro funções interdependentes e iterativas: Govern, que estabelece políticas, responsabilidades e supervisão; Map, que contextualiza o sistema, seus atores, finalidades e impactos; Measure, que analisa e acompanha os riscos; e Manage, que orienta sua priorização, resposta, documentação e monitoramento. O framework prevê revisões periódicas, identificação de riscos emergentes e melhoria contínua. Também estabelece que os riscos mais elevados recebam tratamento prioritário, conforme a tolerância organizacional, as lacunas entre as situações atual e pretendida e o contexto de uso. Seus perfis permitem adaptar os resultados às características de setores, organizações, tecnologias ou aplicações. Essa lógica foi particularizada no perfil para IA generativa, que associa riscos específicos a ações de gestão ao longo do ciclo de vida (Tabassi, 2023; Autio *et al*., 2024).

A ABNT NBR ISO/IEC 23894:2023 apresenta orientação convergente, baseada na estrutura da ABNT NBR ISO 31000. A norma abrange comunicação, definição do contexto, identificação, análise, avaliação, tratamento, monitoramento, análise crítica, registro e relato dos riscos. Para sistemas de IA, essas atividades devem acompanhar os estágios do ciclo de vida e considerar riscos provenientes de dados, modelos, hardware, uso, interações humanas e ambiente sociotécnico, com possíveis consequências para organizações, indivíduos, grupos e sociedade. As medidas de tratamento devem reduzir consequências negativas a níveis aceitáveis e favorecer

resultados positivos, com eficácia verificada e registrada. O monitoramento deve identificar mudanças no contexto, novos riscos e alterações na aceitabilidade dos riscos residuais, formando um ciclo contínuo de identificação, tratamento, verificação e revisão (ABNT, 2023).

A ABNT NBR ISO/IEC 42001:2024 complementa essa perspectiva ao estabelecer requisitos auditáveis para um Sistema de Gestão de Inteligência Artificial, segundo o ciclo Plan–Do–Check–Act. A organização deve definir contexto, partes interessadas, objetivos e riscos; implementar processos e controles; avaliar o desempenho; e promover correções e melhorias. As avaliações de riscos e impactos fundamentam a seleção dos controles, cuja aplicabilidade deve ser justificada. Mudanças relevantes no sistema ou em seu contexto podem exigir nova avaliação, revisão dos controles e atualização documental (ABNT, 2024).

Embora tenham finalidades distintas, os três referenciais convergem na contextualização do risco, na consideração de impactos multidimensionais, na integração ao ciclo de vida, na proporcionalidade das respostas e no monitoramento contínuo. O NIST AI RMF é voluntário e orientado a resultados; a ISO/IEC 23894 orienta o processo de gestão de riscos; e a ISO/IEC 42001 estabelece requisitos para um sistema de gestão. A análise de interoperabilidade da OCDE confirma que referenciais internacionais compartilham componentes como governança, contextualização, avaliação, tratamento, comunicação e acompanhamento, apesar das diferenças terminológicas e estruturais (OECD, 2023). Essa convergência sustenta que um perfil orientado a riscos deve constituir um mecanismo revisável de associação entre cenários de risco, medidas de tratamento e prioridades organizacionais, e não uma lista estática de controles.

2.2 PPSI 2.0, MEDIDAS DE PRIVACIDADE E SEGURANÇA E NECESSIDADE DE SELEÇÃO ORIENTADA A RISCOS

No contexto da Administração Pública Federal, o Programa de Privacidade e Segurança da Informação (PPSI 2.0) orienta a implementação e o acompanhamento de práticas de privacidade, proteção de dados pessoais e segurança da informação. Seu framework reúne 210 medidas sobre governança, gestão de riscos, ativos e dados, configuração segura, acessos, vulnerabilidades, registros de auditoria, desenvolvimento seguro, fornecedores, resposta a incidentes e continuidade, entre outros temas (Brasil, 2026).

As medidas de segurança da informação baseiam-se no CIS Controls v8.1, composto por 153 medidas, 18 controles e três Grupos de Implementação. O GI1 reúne práticas essenciais de higiene cibernética; o GI2 acrescenta medidas adequadas a organizações com maior estrutura e complexidade; e o GI3 incorpora salvaguardas especializadas. O PPSI complementa essa estrutura com medidas de governança, privacidade e proteção de dados pessoais, oferecendo uma base abrangente aos órgãos e entidades da APF (Brasil, 2026).

Embora forneçam uma priorização inicial, os Grupos de Implementação não substituem a análise dos riscos específicos de cada organização ou tecnologia. O guia determina que a adoção das medidas não obrigatórias considere ameaças, vulnerabilidades e impactos, cabendo à organização definir a necessidade e o nível de implementação conforme seu contexto. O PPSI, portanto, não deve ser interpretado como uma lista uniforme aplicável integralmente a qualquer cenário (Brasil, 2026).

Essa seleção orientada a riscos é especialmente relevante para a IA, pois controles gerais podem alcançar apenas parte de determinado cenário. Medidas relacionadas a dados, acessos, desenvolvimento seguro, registros, monitoramento e fornecedores contribuem para tratar riscos de IA, mas não asseguram cobertura suficiente de deriva do modelo, confabulação, ataques adversariais, falta de explicabilidade, automação excessiva ou supervisão humana inadequada. A confiabilidade da IA exige também governança integrada de dados, responsabilidades, transparência, desempenho e prestação de contas (Janssen *et al*., 2020; Mantymaki *et al*., 2022).

A aplicação do PPSI aos riscos de IA requer identificar as medidas compatíveis com as causas, os eventos ou as consequências de cada risco e reconhecer os componentes não alcançados ou dependentes de requisitos específicos. Essa distinção evita tanto desconsiderar controles institucionalizados quanto presumir que práticas gerais de privacidade e segurança cobrem integralmente fenômenos próprios ou intensificados pela IA.

Não se exige, necessariamente, um conjunto independente de controles, mas um perfil adaptativo que preserve a estrutura do PPSI, relacione suas medidas aos riscos identificados e indique as complementações necessárias. Essa proposta converge com o NIST AI RMF, cujos perfis contextualizam resultados e prioridades, e com a ISO/IEC 42001, que exige justificar os controles necessários com base nas avaliações de riscos e impactos (Tabassi, 2023; ABNT, 2024). O perfil atua, assim, como camada de especialização: não transforma o PPSI em framework geral de governança de IA, mas delimita suas contribuições e insuficiências.

2.3 TAXONOMIAS DE RISCOS E CONTRIBUIÇÃO DAS MEDIDAS PARA O TRATAMENTO: COMPATIBILIDADE DOCUMENTAL E CAUSAL

A comparação de riscos extraídos de múltiplos referenciais enfrenta diferenças semânticas anteriores ao mapeamento de controles. Documentos podem empregar termos distintos para fenômenos equivalentes, atribuir sentidos diferentes ao mesmo termo ou reunir, no mesmo nível, fontes de risco, eventos, vulnerabilidades e consequências. Sua simples agregação produz redundâncias e unidades heterogêneas, prejudicando a contagem e a comparação. Portanto, a análise das medidas do PPSI requer previamente uma representação homogênea dos riscos.

Taxonomias organizam objetos de um domínio segundo dimensões e características comuns. Em Sistemas de Informação, atuam como artefatos estruturantes e teorias para análise, pois fornecem conceitos e categorias para representar fenômenos, estabelecer um vocabulário comum e comparar objetos provenientes de fontes distintas (Gregor, 2006; Kundisch *et al*., 2022). Nickerson, Varshney e Muntermann (2013) propõem um método iterativo que combina abordagens conceitual-empírica e empírico-conceitual, partindo da definição do propósito, dos usuários e da característica meta até o atendimento das condições de encerramento. Com base na análise de 164 estudos, Kundisch *et al*. (2022) ampliaram esse método por meio do Extended Taxonomy Design Process e de 26 recomendações sobre definição, construção, avaliação e comunicação. Além de aumentar a transparência e a rastreabilidade, essa abordagem permite atualizar a taxonomia diante de novos riscos ou evidências, aspecto relevante para a evolução da IA.

Neste estudo, a taxonomia exerce função de homogeneização. Os registros extraídos são normalizados como cenários de risco, com preservação da origem documental e distinção, sempre que possível, entre elemento afetado, fonte ou causa, evento e consequência. Termos diferentes são consolidados quando representam o mesmo fenômeno, enquanto riscos relacionados permanecem separados quando apresentam mecanismos causais ou efeitos relevantes distintos. O procedimento produz unidades comparáveis para a análise e evita a inflação artificial da cobertura por duplicidades terminológicas.

A contribuição das medidas para o tratamento dos riscos é avaliada com base em sua compatibilidade causal e documental com cada cenário. A compatibilidade causal ocorre quando

o objeto protegido, o mecanismo ou a finalidade da medida alcança causa, evento ou consequência relevante do cenário. A documental exige que essa relação seja sustentada pelo conteúdo da medida, por seu controle de origem ou pelos referenciais que a fundamentam. Relações baseadas apenas em benefícios genéricos para a segurança ou em interpretações remotas representam menor contribuição para o tratamento. Essa contribuição é classificada como direta, quando a medida alcança explicitamente um elemento central do cenário; parcial, quando cobre apenas parte dele ou requer adaptação; indireta, quando fornece suporte ou condição habilitadora; ou inexistente, quando não há correspondência defensável. Essas categorias expressam a contribuição potencial das medidas para o tratamento dos riscos, e não seu desempenho efetivo.

A contribuição identificada, portanto, não se confunde com eficácia. A primeira é inferida documentalmente e indica potencial lógico de tratamento; a segunda depende da implementação, adequação técnica, responsabilidades, recursos, contexto, monitoramento e resultados. A ISO/IEC 23894 distingue a seleção das opções de tratamento da posterior verificação de sua eficácia (ABNT, 2023). Assim, o mapeamento não comprova redução da probabilidade, do impacto ou do risco residual, nem maturidade ou conformidade organizacional. Ele identifica correspondências auditáveis capazes de subsidiar a seleção, a priorização, a implementação e a avaliação das medidas.

3 METODOLOGIA

3.1 DESENHO DA PESQUISA

A pesquisa adotou abordagem aplicada, documental e predominantemente qualitativa, complementada por estatística descritiva. Seu caráter aplicado decorre da construção de um artefato para verificar em que medida as medidas do PPSI 2.0 contribuem para o tratamento de riscos de IA na Administração Pública Federal (APF). O percurso aproxima-se da ciência do design, que constrói e avalia artefatos para problemas relevantes, embora não tenha sido examinada sua implementação em organizações específicas (Peffers *et al*., 2007). O artefato compreende uma taxonomia normalizada de riscos e uma matriz de correspondência entre riscos e medidas.

A análise documental foi adotada porque as unidades empíricas são enunciados de documentos técnicos, normativos e institucionais. Considerando que esses documentos refletem finalidades, públicos e contextos próprios, a extração preservou a procedência e o contexto dos registros (Bowen, 2009; Dalglish; Khalid; Mcmahon, 2020). A análise qualitativa de conteúdo orientou a interpretação, a decomposição e a comparação dos registros. Complementarmente, frequências absolutas e relativas foram utilizadas para sintetizar o corpus e caracterizar as relações segundo o grau de contribuição para o tratamento, os riscos, as medidas, os controles e as famílias temáticas. Não foram realizados testes de hipótese, e os resultados restringem-se ao material analisado.

A unidade analítica foi o cenário de risco semanticamente delimitado, e não a ocorrência de termos ou as denominações adotadas pelas fontes. Cada cenário deveria ser suficientemente específico para comparação com uma medida de tratamento, em consonância com métodos de construção de taxonomias que recomendam explicitar propósito, característica meta, critérios de agrupamento, iterações e condições de encerramento (Nickerson; Varshney; Muntermann, 2013; Kundish *et al*., 2022).

3.2 CONSTITUIÇÃO DO CORPUS, TRIAGEM E NORMALIZAÇÃO DOS RISCOS

O corpus inicial reuniu 158 registros extraídos de referenciais técnicos e normativos sobre gestão, governança, segurança e riscos de IA, selecionados por sua aplicação institucional e pertinência à APF. Para cada registro, preservaram-se documento, versão ou data, seção ou página, termo original, evidência e tradução ou síntese em português. Esses registros constituíram o corpus de entrada, não 158 riscos independentes, pois as fontes empregavam diferentes vocabulários e níveis de abstração, podendo denominar como risco uma fonte, vulnerabilidade, evento, consequência, propriedade do sistema ou recomendação de tratamento.

A triagem ocorreu em três movimentos. Primeiro, verificaram-se a presença de conteúdo de risco identificável e a suficiência da evidência documental. Treze unidades, como registros de apoio, formulações sem cenário autônomo ou elementos não comparáveis, foram excluídas da taxonomia final, mas tiveram sua linhagem preservada. Em seguida, removeram-se duplicidades literais e consolidaram-se sinônimos explícitos. Por fim, registros de fontes distintas foram reunidos somente quando descreviam o mesmo mecanismo de risco em nível de abstração

compatível. A semelhança temática não justificou a fusão quando havia diferenças relevantes de evento, causa ou consequência.

Para reduzir a heterogeneidade, as unidades foram estruturadas em contexto–fonte–evento–consequência. O contexto identifica o ativo, processo, fase do ciclo de vida ou situação; a fonte corresponde à condição, ao agente, à vulnerabilidade ou à causa; o evento descreve a ocorrência ou mudança de estado; e a consequência representa os possíveis efeitos sobre organizações, pessoas, sociedade, segurança, privacidade, desempenho ou direitos. Quando esses componentes não estavam expressos no documento, foram mantidos vazios ou preenchidos como síntese analítica, distinguindo-se a evidência da fonte da interpretação da pesquisadora.

O processo foi iterativo. A cada consolidação, compararam-se rótulo, descrição, componentes causais e evidências, com retorno às fontes em casos ambíguos. A característica meta foi definida como “riscos associados ao desenvolvimento, disponibilização ou uso de IA que possam demandar tratamento por mecanismos institucionais de governança, privacidade ou segurança da informação”. As iterações terminaram quando todos os registros elegíveis estavam alocados, não surgiam novos cenários ou dimensões materiais, não havia duplicidades conceituais identificáveis e cada risco possuía denominação e descrição distinguíveis. O processo resultou em 87 riscos normalizados. A taxonomia homogeneizou o corpus, sem representar prevalência ou criticidade dos riscos.

3.3 MAPEAMENTO N:N ENTRE RISCOS E MEDIDAS DO PPSI 2.0

Os 87 riscos foram confrontados com as 210 medidas do PPSI 2.0 por meio de um relacionamento muitos-para-muitos (N:N), pois um risco pode exigir várias medidas e uma medida pode atuar sobre diferentes riscos. A unidade de codificação foi o par risco–medida. Para cada relação material identificada, registraram-se os identificadores e as denominações do risco e da medida, a natureza do risco, o texto da medida, o controle correspondente do PPSI, o grau de contribuição para o tratamento, a justificativa técnica e, quando aplicável, o tema de fortalecimento.

A contribuição de cada medida para o tratamento de riscos foi avaliada com base em três critérios: se o objeto protegido pela medida correspondia a elemento relevante do cenário; se seu mecanismo alcançava a fonte, o evento ou a consequência; e se a relação era sustentada pelo

texto da medida, por seu resumo ou pelo controle correspondente. Somente foram registradas associações com fundamento documental e causal defensável. Benefícios genéricos de segurança não foram suficientes. Quando a medida alcançava vários componentes do mesmo risco, atribuiu-se ao par um único grau, correspondente à relação predominante, evitando dupla contagem.

A contribuição para o tratamento foi classificada como:

- direta, quando a medida atua explicitamente sobre elemento central do risco e pode ser aplicada sem requisito adicional específico de IA;
- parcial, quando alcança componente relevante, mas não todo o mecanismo, ou requer adaptação, detalhamento ou complemento próprio de IA;
- indireta, quando fornece condição habilitadora, como governança, papéis, inventário, documentação ou capacidade de resposta, sem tratar o mecanismo principal;
- inexistente, quando não há compatibilidade causal e documental suficiente.

Essas categorias expressam a contribuição potencial das medidas para o tratamento dos riscos, não sua prioridade, criticidade ou eficácia operacional. O confronto compreendeu 18.270 pares possíveis (87 × 210). Para manter a matriz operacional e auditável, apenas as relações classificadas como diretas, parciais ou indiretas foram registradas individualmente. Os demais pares constituíram o conjunto residual de contribuição inexistente. Ao final, foram identificadas 921 relações materiais, das quais 183 foram classificadas como diretas, 607 como parciais e 131 como indiretas.

3.4 SÍNTESE, TRILHA DE AUDITORIA E LIMITAÇÕES DE VALIDADE

A síntese quantitativa considerou a distribuição das relações por grau, as quantidades de medidas por risco e de riscos por medida, a existência de relação direta e a distribuição das lacunas por família temática. A ponderação adotada na planilha — 3 para direta, 2 para parcial e 1 para indireta — serviu apenas à ordenação e agregação descritiva. Não constitui escala intervalar validada nem medida de eficácia, redução do risco residual ou prioridade de implantação. Relações inexistentes não receberam peso negativo.

A trilha de auditoria foi mantida em duas estruturas. A primeira preservou a relação entre registros originais e riscos normalizados, incluindo fonte, localização, evidência, componentes do cenário e decisões de consolidação ou exclusão. A segunda registrou os pares risco–medida,

seus identificadores, graus e justificativas, além de uma tabela-mestre das 210 medidas e seus totais. Também foram conservadas versões intermediárias e regras de codificação, permitindo reconstruir o percurso entre fonte e classificação final. Essa trilha favorece transparência, dependabilidade e confirmabilidade da análise qualitativa (Nowell *et al*., 2017).

Quanto à validade de construto, a decomposição dos cenários de risco e a classificação das contribuições das medidas como diretas, parciais, indiretas ou inexistentes dependem de escolhas analíticas, sobretudo diante da amplitude de algumas medidas do PPSI. A preservação das evidências e o registro das justificativas tornam o processo auditável e reduzem, embora não eliminem, a possibilidade de interpretações alternativas. Quanto à validade das inferências, a contribuição identificada com base na compatibilidade documental e causal indica apenas uma possibilidade lógica de tratamento; não constitui evidência de implementação, eficácia, redução da probabilidade ou do impacto, nem de alteração do risco residual. Quanto à validade externa, os resultados limitam-se ao corpus selecionado, à versão analisada do PPSI 2.0 e ao contexto da APF, podendo demandar revisão diante da atualização dos referenciais ou de sua aplicação a outros setores.

A codificação foi realizada por uma única pesquisadora, sem painel de especialistas ou mensuração de concordância interavaliadores. Embora o livro de códigos, as regras de decisão, os identificadores e a trilha de auditoria tenham atuado como salvaguardas, não substituem validação independente (O'Connor; Joffe, 2020). Estudos futuros poderão submeter amostras dos pares à codificação independente, avaliar a concordância, discutir divergências e validar a utilidade do mapeamento com especialistas e órgãos da APF.

4. RESULTADOS

4.1 NORMALIZAÇÃO DOS REGISTROS E CONSTITUIÇÃO DA LISTA DE RISCOS

O corpus inicial reuniu 158 registros dos referenciais selecionados, com diferentes níveis de abstração e unidades conceituais, incluindo riscos, vulnerabilidades, fontes, eventos, consequências, exemplos e recomendações de tratamento. Também foram encontrados termos distintos para fenômenos equivalentes ou parcialmente sobrepostos.

A triagem e a normalização resultaram em 87 cenários de risco, estruturados por contexto, fonte ou causa, evento e consequência. O processo envolveu consolidação de equivalências conceituais, desagregação de registros com múltiplos eventos e padronização de formulações genéricas. Registros foram reunidos quando descreviam o mesmo mecanismo de risco e mantidos separados quando apresentavam diferenças materiais de causa, evento ou consequência.

Outras 13 unidades foram excluídas da lista principal por representarem oportunidades de uso da IA, práticas de tratamento, capacidades organizacionais, exemplos informativos ou formulações sem evento de risco delimitado. Essas exclusões, assim como as consolidações e desagregações, preservaram a identificação, a fonte, a localização e a justificativa da decisão. Os quantitativos da Tabela 1 não constituem etapas mutuamente excludentes, pois vários registros podem originar um risco, enquanto um registro composto pode gerar mais de um cenário.

Tabela 1. Síntese da normalização dos riscos de IA

Indicador	**Resultado**	**Significado**
Registros no corpus inicial	158	Unidades extraídas dos documentos analisados
Riscos normalizados	87	Cenários comparáveis submetidos ao mapeamento
Unidades excluídas da lista principal	13	Oportunidades, controles, capacidades ou formulações sem evento de risco delimitado
Estrutura de normalização	Contexto-fonte-evento-consequência	Padrão empregado para homogeneizar os cenários
Linhagem documental	Preservada	Manutenção da fonte, localização e decisão de tratamento de cada registro

Fonte: elaboração própria com base na taxonomia normalizada.

A lista de 87 riscos constituiu, portanto, uma camada analítica sobre o corpus original, evitando duplicidades decorrentes de diferenças terminológicas e fornecendo unidades comparáveis para o mapeamento com as medidas do PPSI.

4.2 CONTRIBUIÇÃO DAS MEDIDAS DO PPSI 2.0 PARA O TRATAMENTO DOS RISCOS DE IA

Os 87 riscos foram confrontados com as 210 medidas do PPSI 2.0 em uma estrutura muitos-para-muitos, totalizando 18.270 combinações possíveis. A matriz registrou somente as

921 relações com compatibilidade documental ou causal suficiente para classificação como direta, parcial ou indireta.

Tabela 2. Indicadores da análise de contribuição das medidas do PPSI 2.0

Indicador	Quantidade	Percentual
Riscos normalizados analisados	87	100%
Medidas do PPSI 2.0 analisadas	210	100%
Combinações possíveis	18.270	-
Relações materiais identificadas	921	100% das relações
Relações diretas	183	19,9%
Relações parciais	607	65,9%
Relações indiretas	131	14,2%
Riscos com ao menos uma relação direta	35	40,2% dos riscos
Riscos sem relação direta, mas com relação parcial	52	59,8% dos riscos
Medidas com alguma relação material	158	75,2% das medidas
Medidas sem relação material identificada	52	24,8% das medidas

Fonte: elaboração própria com base na matriz de risco-medida.

A predominância das relações parciais indica que o PPSI alcança componentes relevantes dos riscos, mas, em geral, não contempla integralmente seus mecanismos nem estabelece requisitos específicos para IA. Os 35 riscos com ao menos uma relação direta concentraram-se em temas já abrangidos por mecanismos expressos de segurança e privacidade, como proteção de dados pessoais, identidades e acessos, desenvolvimento seguro, vulnerabilidades, fornecedores, auditoria e resposta a incidentes.

Os outros 52 riscos apresentaram apenas relações parciais ou indiretas. Predominaram nesse grupo fenômenos relacionados ao comportamento dos modelos, à avaliação de resultados e à interação humano–IA. Embora medidas de gestão de riscos, logs, testes, mudanças e governança contribuam para seu tratamento, elas não definem requisitos específicos, como métricas de desempenho, limites de variação, detecção de deriva, testes adversariais, explicabilidade e supervisão humana qualificada.

A deriva de modelo exemplifica essa limitação. Logs e gestão de mudanças fornecem informações e processos de apoio, mas não estabelecem mecanismos para detectar, compreender e responder a alterações nos dados ou na relação entre entradas e resultados, atividades próprias da gestão de deriva conceitual (Lu *et al.*, 2019).

Das 210 medidas, 158 apresentaram alguma relação material e 52 não foram associadas aos riscos analisados. Essa ausência não indica que sejam dispensáveis, mas apenas que não foi identificada ligação causal suficientemente defensável no recorte adotado. Da mesma forma, a

quantidade de relações não representa eficácia, prioridade ou capacidade de redução do risco residual.

4.3 FAMÍLIAS DE LACUNAS E NECESSIDADES DE FORTALECIMENTO

A análise dos componentes não alcançados pelas medidas e das justificativas das relações parciais permitiu agrupar as necessidades de complementação em oito famílias. Elas não constituem novos riscos, controles prontos ou ranking de criticidade, mas indicam temas nos quais um perfil complementar de IA poderia estabelecer requisitos e evidências adicionais.

Tabela 3. Famílias de fortalecimento identificadas

Prioridade de fortalecimento	**Família**	**Índice de lacuna**
Alta	Validação, desempenho e monitoramento de modelos	46,7%
Alta	Segurança de IA e testes adversariais	32,5%
Alta	Direitos, transparência, explicabilidade e supervisão humana	30,0%
Alta	Competências, cultura e fatores humanos	30,0%
Média	Governança, política, papéis e inventário de IA	23,1%
Média	Terceiros, cadeia de IA, infraestrutura e sustentabilidade	20,0%
Média	Incidentes, comunicação e contestabilidade	20,0%
Baixa	Qualidade e proveniência de dados de IA	16,7%

Fonte: elaboração própria com base na matriz de risco-medida. Nota: o índice expressa necessidade de fortalecimento do perfil e não criticidade do risco, prioridade de implantação ou risco residual.

A maior lacuna ocorreu em validação, desempenho e monitoramento de modelos (46,7%). Embora o PPSI contemple testes, mudanças, auditoria e monitoramento de eventos, não estabelece critérios específicos para validação prévia, métricas e limites de desempenho, comparação entre grupos ou contextos, detecção de degradação e deriva, revalidação e suspensão ou retirada de modelos. Essas necessidades correspondem às atividades de detecção, compreensão e adaptação reconhecidas na gestão de deriva conceitual (Lu *et al*., 2019).

Segurança de IA e testes adversariais apresentou índice de 32,5%. As medidas de desenvolvimento seguro, gestão de vulnerabilidades, testes de intrusão e modelagem de ameaças oferecem cobertura relevante, mas não especificam testes para manipulação dos dados de treinamento, evasão, envenenamento, extração ou inversão de modelos e entradas adversariais. Esses ataques exploram o processo de aprendizagem e exigem avaliações além das aplicadas a sistemas convencionais (Biggio; Roli, 2018).

Direitos, transparência, explicabilidade e supervisão humana (30,0%) reuniu lacunas relativas à comunicação de limitações, explicação de resultados, identificação das pessoas afetadas, avaliação de impactos, revisão humana e prevenção da confiança excessiva na automação. As medidas do PPSI relacionadas à transparência e à proteção de dados não alcançam integralmente impactos produzidos por inferências ou decisões do modelo. Além disso, a supervisão humana depende de autoridade, tempo, informação e competência para contestar resultados automatizados (Green, 2022).

Competências, cultura e fatores humanos também apresentou índice de 30,0%. As medidas de capacitação não incluem conteúdos específicos sobre limitações, vieses, interpretação de resultados, automação excessiva, usos permitidos e responsabilidades no ciclo de vida da IA. Conhecimento, competências, recursos e alinhamento organizacional são condições relevantes para a adoção responsável da tecnologia (Johnk; Weissert; Wyrtki, 2021).

As demais famílias apresentaram índices entre 16,7% e 23,1%. Em governança, destacaram-se a ausência de inventário sistemático de sistemas e modelos e a definição insuficiente de responsáveis, finalidades e níveis de risco. Em terceiros, infraestrutura e sustentabilidade, faltaram requisitos sobre procedência de modelos e dados, dependências e impactos ambientais. Em incidentes e contestabilidade, observaram-se lacunas na definição de incidentes algorítmicos, comunicação às pessoas afetadas e mecanismos de contestação. Em qualidade e proveniência de dados, apesar das medidas existentes de inventário, proteção e retenção, permaneceram lacunas relacionadas à adequação estatística, representatividade, rotulagem, rastreabilidade e documentação dos dados de treinamento e avaliação.

Em síntese, o PPSI oferece pontos de apoio para as oito famílias, mas sua cobertura é desigual. As relações diretas concentram-se em privacidade e segurança tradicionais, enquanto riscos ligados ao funcionamento, à avaliação e ao uso sociotécnico da IA permanecem majoritariamente cobertos de forma parcial. Os resultados justificam o fortalecimento do framework, mas não permitem inferir eficácia operacional, maturidade institucional ou redução do risco residual. A operacionalização da governança de IA requer requisitos, processos, responsabilidades e instrumentos técnicos específicos, integrados às estruturas existentes (Mantymaki *et al*., 2022).

5 DISCUSSÃO

5.1 O PPSI 2.0 COMO BASELINE PARA OS RISCOS DE IA

Os resultados sustentam que o PPSI 2.0 constitui uma baseline útil para o tratamento dos riscos de IA, mas não um framework completo de governança dessa tecnologia. Das 210 medidas analisadas, 158 apresentaram alguma relação material com os riscos, e todos os 87 cenários normalizados possuíam ao menos uma medida direta ou parcial. Esse resultado mostra que a introdução da IA não torna obsoletos os mecanismos existentes de privacidade e segurança da informação. Sistemas de IA continuam dependentes de dados, identidades, códigos, aplicações, infraestrutura, registros, fornecedores e processos de resposta a incidentes, objetos já alcançados por diferentes segmentos do PPSI.

Nesse contexto, a utilidade do PPSI decorre, portanto, de sua capacidade de tratar componentes subjacentes aos riscos de IA. Medidas de gestão de acessos podem reduzir a possibilidade de manipulação não autorizada de modelos e dados; medidas de desenvolvimento seguro e gestão de vulnerabilidades podem atuar sobre falhas na implementação; medidas de proteção de dados podem limitar consequências sobre titulares; e requisitos relacionados a fornecedores, registros e incidentes podem fortalecer rastreabilidade, responsabilização e resposta. Isso não significa que essas medidas tenham sido concebidas para riscos algorítmicos, mas que seus mecanismos de atuação permanecem relevantes quando incorporados ao ciclo de vida da IA.

Essa interpretação também encontra respaldo na literatura sobre governança organizacional de IA. Mäntymäki *et al*. (2022) posicionam a governança de IA como parte de uma estrutura organizacional na qual já coexistem governança corporativa, governança de TI e governança de dados. Assim, governar a IA não significa necessariamente instituir uma estrutura paralela e independente, mas especializar capacidades existentes e coordená-las com novos requisitos.

Para a APF, essa abordagem oferece uma vantagem institucional: os órgãos podem partir de um instrumento já inserido em seu ambiente normativo e operacional, em vez de criar um sistema de controles inteiramente separado. Essa integração é especialmente relevante porque os riscos da IA no setor público não se limitam à segurança técnica, alcançando prestação de contas,

direitos, transparência, legitimidade das decisões e qualidade dos serviços públicos (Busuioc, 2021; Zuiderwijk; Chen; Salem, 2021). O PPSI funciona, assim, como ponto de partida para a dimensão de privacidade e segurança, enquanto um perfil complementar acrescentaria requisitos relacionados ao comportamento, à avaliação e ao uso sociotécnico da IA.

A conclusão correta não é, portanto, que o PPSI "já cobre os riscos de IA", mas que ele fornece uma infraestrutura de controles aproveitável. Essa formulação preserva a finalidade original do programa e, simultaneamente, justifica sua adaptação. A contribuição identificada demonstra uma interseção material entre os dois domínios, sem transformar o PPSI em um framework geral de governança de IA.

5.2 SIGNIFICADO DA PREDOMINÂNCIA DE RELAÇÕES PARCIAIS

A predominância de relações parciais constitui o achado mais relevante da análise. Das 921 relações materiais, 607 foram classificadas como parciais, correspondendo a 65,9% do total. Além disso, 52 dos 87 riscos, ou 59,8%, não apresentaram nenhuma medida diretamente aplicável. Esses números mostram que a maior parte da cobertura não é inexistente, mas também não pode ser considerada suficiente apenas pela presença de alguma associação.

A relação parcial significa que a medida alcança um componente causal, um objeto protegido ou uma consequência do cenário, mas não contempla todo o mecanismo do risco ou não contém requisitos específicos para IA. Medidas de logs, por exemplo, fornecem dados necessários ao monitoramento, mas não definem métricas, limites de desempenho ou procedimentos para detectar e tratar deriva do modelo. Testes de segurança de aplicações podem identificar vulnerabilidades convencionais, mas não necessariamente avaliam envenenamento de dados, evasão, extração ou inversão de modelos. Da mesma forma, medidas gerais de transparência e capacitação não garantem explicações apropriadas sobre resultados algorítmicos nem supervisão humana efetiva.

A classificação gradual evita dois erros opostos. O primeiro seria descartar o PPSI por não ter sido criado especificamente para IA, desconsiderando controles que podem contribuir para o tratamento dos riscos. O segundo seria equiparar qualquer relação a cobertura integral. Se as relações diretas, parciais e indiretas fossem agregadas em uma categoria binária de "aderente", a matriz produziria uma percepção inflada de suficiência. A literatura sobre operacionalização

da governança de IA alerta justamente para a distância entre princípios ou instrumentos gerais e os mecanismos concretos necessários para colocá-los em prática (Morley *et al*., 2020).

Também não se pode presumir que várias medidas parciais, quando combinadas, produzam automaticamente cobertura integral. Essa possibilidade precisa ser examinada em cada cenário, verificando se os diferentes mecanismos são complementares e se permanecem componentes sem tratamento. Inversamente, a existência de uma relação direta não demonstra que o risco esteja integralmente coberto, pois outras causas ou consequências podem exigir medidas adicionais.

A predominância das relações parciais deve ser interpretada, portanto, como evidência de uma arquitetura incompleta, mas aproveitável. Ela indica que o PPSI contém mecanismos pertinentes, ao mesmo tempo que revela a necessidade de especificá-los para o ciclo de vida da IA. Trata-se do principal fundamento empírico para um perfil complementar: não substituir as 210 medidas, mas identificar, para cada grupo de riscos, quais medidas podem ser reutilizadas, quais dependem de adaptação e quais requisitos ainda precisam ser acrescentados.

5.3 COMPLEMENTAÇÕES JUSTIFICADAS PELOS RESULTADOS

As oito famílias identificadas orientam o fortalecimento do framework, mas não representam um ranking de riscos organizacionais. A maior lacuna ocorreu em validação, desempenho e monitoramento de modelos (46,7%). Um perfil complementar deve prever documentação da finalidade e dos limites de uso, validação prévia, métricas de desempenho, comparação entre grupos ou contextos, limites de tolerância, monitoramento de degradação e deriva, revalidação após mudanças e critérios de suspensão ou retirada. A deriva exige atividades próprias de detecção, compreensão e adaptação, não se confundindo com a manutenção de logs (LU *et al*., 2019).

Segurança de IA e testes adversariais apresentou lacuna de 32,5%. As medidas de desenvolvimento seguro, testes de intrusão e gestão de vulnerabilidades devem ser mantidas, mas complementadas por modelos de ameaça específicos para IA. Conforme a tecnologia e o contexto, devem abranger manipulação de dados de treinamento, evasão, envenenamento, extração e inversão de modelos, robustez e dependências externas. Esses ataques exploram

características do aprendizado e exigem avaliações além da segurança convencional de aplicações (Biggio; Roli, 2018).

Direitos, transparência, explicabilidade e supervisão humana (30,0%) requerem medidas que alcancem tanto o tratamento dos dados quanto os resultados do sistema. Isso inclui avaliar impactos sobre pessoas e grupos, comunicar limitações e o uso de IA, registrar decisões, fornecer explicações adequadas e assegurar contestação e revisão. Na APF, a opacidade pode comprometer a responsabilização e a contestação de decisões estatais (Busuioc, 2021). A presença humana, isoladamente, é insuficiente: a supervisão exige competência, informação, tempo, autoridade e condições institucionais para contrariar resultados algorítmicos (Green, 2022).

Competências, cultura e fatores humanos também apresentou lacuna de 30,0%. As capacitações devem ser diferenciadas por papel e abranger gestores, equipes técnicas, áreas de negócio, compradores públicos, operadores, auditores e revisores humanos. Os conteúdos devem tratar de limitações dos modelos, interpretação dos resultados, vieses, automação excessiva, usos permitidos, resposta a falhas e responsabilidades no ciclo de vida. Conhecimento, recursos, alinhamento estratégico e competências influenciam a prontidão organizacional para adoção da IA (Johnk; Weissert; Wyrtki, 2021).

As demais famílias completam essa especialização. Governança, política, papéis e inventário devem identificar sistemas, modelos, responsáveis, finalidades e contextos de uso. Qualidade e proveniência de dados exigem registros de origem, composição, transformações, adequação e restrições. Terceiros, infraestrutura e sustentabilidade requerem cláusulas e evidências sobre modelos adquiridos, dependências, atualizações e responsabilidades dos fornecedores. Incidentes, comunicação e contestabilidade devem estabelecer critérios para reconhecer incidentes algorítmicos, interromper sistemas, comunicar pessoas afetadas e registrar correções.

Esses requisitos podem compor uma camada associada às medidas do PPSI, sem formar catálogo independente, indicando o componente não coberto, o requisito adicional, as evidências esperadas e os eventos de revisão. Essa configuração reduz duplicidades, preserva a arquitetura institucional e transforma a matriz de risco-medida em instrumento de seleção e adaptação.

5.4 ALCANCE E LIMITES DAS CONCLUSÕES

A análise demonstra cobertura potencial em termos documentais e causais, mas não a eficácia das medidas, pois não examinou sua implementação, funcionamento ou capacidade de reduzir probabilidade e impacto. Mesmo medidas diretamente aplicáveis podem ser ineficazes quando executadas sem recursos, responsabilidades, monitoramento ou adequação ao contexto.

O estudo também não avalia maturidade institucional, o que exigiria evidências sobre formalização, implementação, recursos, competências, supervisão e melhoria contínua em órgãos específicos. A existência de uma medida no PPSI ou de uma relação na matriz não comprova capacidade organizacional para executá-la.

Tampouco foi estimado o risco residual, pois isso dependeria do risco inerente, da probabilidade, do impacto, da eficácia dos controles existentes e das condições remanescentes após o tratamento. A quantidade de medidas relacionadas não substitui a avaliação de eficácia: várias medidas frágeis podem reduzir menos risco que uma medida central adequadamente implementada.

Os resultados também não definem prioridades para órgãos específicos. A classificação das famílias em alta, média ou baixa representa necessidade relativa de fortalecimento do perfil, não criticidade dos riscos ou ordem obrigatória de implantação. A priorização depende da missão, dos sistemas e dados envolvidos, das pessoas afetadas, das obrigações legais, da criticidade dos serviços, da probabilidade, do impacto, do apetite ao risco e dos recursos disponíveis.

Por fim, o caráter documental e interpretativo da pesquisa, conduzida por uma única pesquisadora, limita a confiabilidade das classificações. Segundo codificador, painel de especialistas ou aplicação empírica poderiam validá-las ou refiná-las. Diante da necessidade de estudos empíricos sobre implementação, desempenho e impactos da IA no setor público (Zuiderwijk; Chen; Salem, 2021), a contribuição deste estudo situa-se em etapa anterior: estabelecer uma ponte rastreável entre riscos de IA e medidas existentes, indicando possibilidades de tratamento e lacunas que exigem complementação.

6 CONCLUSÃO

Este estudo avaliou a contribuição das medidas do PPSI 2.0 no tratamento dos riscos de inteligência artificial relevantes para a Administração Pública Federal e identificou os domínios que exigem complementação. Os resultados indicam que o PPSI constitui uma baseline institucional útil, porém insuficiente para o tratamento integral desses riscos. Entre os 87 riscos normalizados e as 210 medidas, foram identificadas 921 relações materiais: 183 diretas, 607 parciais e 131 indiretas. Apenas 35 riscos apresentaram ao menos uma medida diretamente aplicável, enquanto 52 permaneceram sem cobertura direta. A predominância de relações parciais (65,9%) demonstra que o PPSI alcança componentes relacionados à privacidade, segurança cibernética, acessos, desenvolvimento seguro, terceiros e resposta a incidentes, mas não contempla integralmente fenômenos próprios do comportamento e do ciclo de vida da IA.

A principal contribuição do estudo é a construção de uma ponte rastreável entre riscos de IA e medidas institucionais da APF. Para isso, 158 registros documentais foram normalizados em 87 cenários comparáveis e relacionados às medidas por meio de um mapeamento muitos-para-muitos, com contribuições classificadas como diretas, parciais, indiretas ou inexistentes. Essa gradação preserva a contribuição de medidas gerais sem presumir cobertura suficiente, enquanto a linhagem documental e as justificativas tornam o processo revisável.

Os resultados sustentam uma arquitetura em duas camadas: o PPSI como baseline de privacidade e segurança da informação e um perfil complementar para IA. Sem substituir o catálogo existente, esse perfil deve especificar requisitos, responsabilidades e evidências adicionais para validação e monitoramento de modelos, segurança adversarial, direitos e transparência, supervisão humana e competências. Essa integração aproveita estruturas organizacionais existentes, evita duplicidades e contribui para converter princípios de governança de IA em mecanismos operacionalizáveis (Mantymaki *et al*., 2022; Morley *et al*., 2020).

As conclusões estão limitadas ao caráter documental e predominantemente qualitativo da pesquisa, conduzida por uma única pesquisadora, sem segundo codificador, validação por especialistas ou aplicação em órgão específico. As classificações expressam compatibilidade causal e documental potencial, não eficácia das medidas, maturidade institucional, redução da

probabilidade ou do impacto, risco residual ou prioridade organizacional. A quantidade de relações também não representa a importância dos riscos nem o desempenho dos controles.

Trabalhos futuros devem validar a taxonomia e a matriz com especialistas e múltiplos codificadores, avaliar a concordância entre julgamentos e realizar aplicações-piloto em órgãos da APF. Recomenda-se ainda investigar a implementação e a eficácia das medidas, incorporar probabilidade, impacto e risco residual, converter as oito famílias de fortalecimento em requisitos verificáveis e acompanhar longitudinalmente mudanças nos sistemas, riscos e referenciais regulatórios. Essa agenda responde à necessidade de ampliar estudos empíricos sobre os impactos da IA na governança pública (Zuiderwijk; Chen; Salem, 2021).

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO/IEC 23894:2023: tecnologia da informação — inteligência artificial — orientações sobre gestão de riscos. Rio de Janeiro: ABNT, 2023.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO/IEC 42001:2024: tecnologia da informação — inteligência artificial — sistema de gestão. Rio de Janeiro: ABNT, 2024.

AUTIO, Chloe *et al*. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. Gaithersburg: **National Institute of Standards and Technology**, 2024. (NIST AI 600-1). DOI: https://doi.org/10.6028/NIST.AI.600-1.

BIGGIO, Battista; ROLI, Fabio. Wild patterns: ten years after the rise of adversarial machine learning. ***Pattern Recognition***, v. 84, p. 317–331, 2018. DOI: https://doi.org/10.1016/j.patcog.2018.07.023.

BOWEN, Glenn A. Document analysis as a qualitative research method. **Qualitative Research Journal**, v. 9, n. 2, p. 27–40, 2009. DOI: https://doi.org/10.3316/QRJ0902027.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. Inteligência artificial. Brasília, DF: **MGI**, 2026a. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1. Acesso em: 22 jul. 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. Guia do Framework de Privacidade e Segurança da Informação: Programa de Privacidade e Segurança da Informação (PPSI 2.0). Versão 1.2. Brasília, DF: **MGI**, 2026b. Disponível em: https://www.gov.br/governodigital/pt-br/privacidade-e-seguranca/ppsi-2.0/arquivos/guiafwppsi2-0-v1-2-11062026.pdf. Acesso em: 22 jul. 2026.

BUSUIOC, Madalina. Accountable artificial intelligence: holding algorithms to account. **Public Administration Review**, v. 81, n. 5, p. 825–836, 2021. DOI: https://doi.org/10.1111/puar.13293.

DALGLISH, Sarah L.; KHALID, Hina; MCMAHON, Shannon A. Document analysis in health policy research: the READ approach. Health Policy and Planning, v. 35, n. 10, p. 1424–1431, 2020. DOI: https://doi.org/10.1093/heapol/czaa064.

GREEN, Ben. The flaws of policies requiring human oversight of government algorithms. **Computer Law & Security Review**, v. 45, art. 105681, 2022. DOI: https://doi.org/10.1016/j.clsr.2022.105681.

GREGOR, Shirley. The nature of theory in information systems. **MIS Quarterly**, v. 30, n. 3, p. 611–642, 2006. DOI: https://doi.org/10.2307/25148742.

JANSSEN, Marijn; BROUS, Paul; ESTEVEZ, Elsa; BARBOSA, Luis S.; JANOWSKI, Tomasz. Data governance: organizing data for trustworthy artificial intelligence. **Government**

Information Quarterly, v. 37, n. 3, art. 101493, 2020. DOI: https://doi.org/10.1016/j.giq.2020.101493.

JÖHNK, Jan; WEISSERT, Malte; WYRTKI, Katrin. Ready or not, AI comes: an interview study of organizational AI readiness factors. **Business & Information Systems Engineering,** v. 63, n. 1, p. 5–20, 2021. DOI: https://doi.org/10.1007/s12599-020-00676-7.

KUNDISCH, Dennis; MUNTERMANN, Jan; OBERLÄNDER, Anna Maria; RAU, Daniel; RÖGLINGER, Maximilian; SCHOORMANN, Thorsten; SZOPINSKI, Daniel. An update for taxonomy designers: methodological guidance from information systems research. **Business & Information Systems Engineering**, v. 64, n. 4, p. 421–439, 2022. DOI: https://doi.org/10.1007/s12599-021-00723-x.

LU, Jie; LIU, Anjin; DONG, Fan; GU, Feng; GAMA, João; ZHANG, Guangquan. Learning under concept drift: a review. ***IEEE Transactions on Knowledge and Data Engineering***, v. 31, n. 12, p. 2346–2363, 2019. DOI: https://doi.org/10.1109/TKDE.2018.2876857.

MÄNTYMÄKI, Matti; MINKKINEN, Matti; BORG, Anna; RIEKKI, Jukka. Defining organizational AI governance. **AI and Ethics**, v. 2, p. 603–609, 2022. DOI: https://doi.org/10.1007/s43681-022-00143-x.

MORLEY, Jessica; FLORIDI, Luciano; KINSEY, Libby; ELHALAL, Anat. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. **Science and Engineering Ethics**, v. 26, n. 4, p. 2141–2168, 2020. DOI: https://doi.org/10.1007/s11948-019-00165-5.

NICKERSON, Robert C.; VARSHNEY, Upkar; MUNTERMANN, Jan. A method for taxonomy development and its application in information systems. **European Journal of Information Systems**, v. 22, n. 3, p. 336–359, 2013. DOI: https://doi.org/10.1057/ejis.2012.26.

NOWELL, Lorelli S.; NORRIS, Jill M.; WHITE, Deborah E.; MOULES, Nancy J. Thematic analysis: striving to meet the trustworthiness criteria. **International Journal of Qualitative Methods**, v. 16, p. 1–13, 2017. DOI: https://doi.org/10.1177/1609406917733847.

O'CONNOR, Cliodhna; JOFFE, Helene. Intercoder reliability in qualitative research: debates and practical guidelines. **International Journal of Qualitative Methods**, v. 19, p. 1–13, 2020. DOI: https://doi.org/10.1177/1609406919899220.

OECD. Common guideposts to promote interoperability in AI risk management. Paris: OECD Publishing, 2023. (**OECD Artificial Intelligence Papers**, n. 5). DOI: https://doi.org/10.1787/ba602d18-en.

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A.; CHATTERJEE, Samir. A design science research methodology for information systems research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45–77, 2007. DOI: https://doi.org/10.2753/MIS0742-1222240302.

SUN, Tara Qian; MEDAGLIA, Rony. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. **Government Information Quarterly**, v. 36, n. 2, p. 368–383, 2019. DOI: https://doi.org/10.1016/j.giq.2018.09.008.

TABASSI, Elham. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg: **National Institute of Standards and Technology**, 2023. (NIST AI 100-1). DOI: https://doi.org/10.6028/NIST.AI.100-1.

WIRTZ, Bernd W.; WEYERER, Jan C.; GEYER, Carolin. Artificial intelligence and the public sector: applications and challenges. **International Journal of Public Administration**, v. 42, n. 7, p. 596–615, 2019. DOI: https://doi.org/10.1080/01900692.2018.1498103.

ZUIDERWIJK, Anneke; CHEN, Yu-Che; SALEM, Fadi. Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda. **Government Information Quarterly**, v. 38, n. 3, art. 101577, 2021. DOI: https://doi.org/10.1016/j.giq.2021.101577.