ÁREA TEMÁTICA: Tecnologia, Inteligência Artificial e Transformação Digital em Administração

CATEGORIA: ENANGRAD PLENO

SHADOW AI NO SETOR PÚBLICO: RISCOS À INTEGRIDADE INFORMACIONAL E À *ACCOUNTABILITY*

Resumo: O uso de inteligência artificial generativa por agentes públicos, por iniciativa própria e à margem das estruturas formais de governança, fenômeno denominado *shadow AI*, difundiu-se mais rapidamente que a capacidade institucional de disciplinálo. O estudo analisou os riscos institucionais desse uso informal para a integridade informacional e para a *accountability* dos processos decisórios. Adotou-se pesquisa qualitativa e exploratório-descritiva, por revisão integrativa nas bases Scopus, Web of Science e SciELO, com análise de conteúdo categorial de 21 estudos. Os achados apontam adoção superior à maturidade dos arranjos institucionais e riscos em dois eixos: fragilização da fidedignidade pela alucinação dos modelos, apagamento da rastreabilidade e exposição de dados sob sigilo; e indefinição sobre quem responde e perante qual instância. Conclui-se que o fenômeno rompe a cadeia documental da racionalidade burocrática. A contribuição teórica está em articular a teoria da burocracia, a tradição arquivística e a *accountability* algorítmica, mostrando que a integridade informacional condiciona materialmente a *accountability*, com subsídios a gestores e ao controle.

Palavras-Chave: *Shadow AI*; Inteligência artificial generativa; Administração pública; Integridade informacional; *Accountability*.

Abstract: The use of generative artificial intelligence by public servants on their own initiative, outside formal governance structures, a phenomenon known as shadow AI, has spread faster than the institutional capacity to regulate it. This study analyzed the institutional risks of such informal use for informational integrity and for the accountability of decision-making processes. A qualitative, exploratory-descriptive integrative review was conducted in the Scopus, Web of Science, and SciELO databases, with categorical content analysis of 21 studies. The findings show adoption outpacing the maturity of institutional arrangements and risks along two axes: weakened reliability from model hallucination, erasure of traceability, and exposure of legally protected data; and uncertainty over who answers and before which forum. It is concluded that the phenomenon breaks the documentary chain sustaining bureaucratic rationality. The theoretical contribution lies in articulating bureaucratic theory, the archival tradition, and algorithmic accountability, showing that informational integrity is a material condition for accountability, with input to managers and oversight bodies.

Keywords: Shadow AI; Generative artificial intelligence; Public administration; Informational integrity; Accountability.

1. Introdução

A incorporação de tecnologias digitais à administração pública tem redefinido as rotinas dos agentes públicos e os modos de produção das decisões administrativas. Nesse movimento, a inteligência artificial generativa, sustentada por grandes modelos de linguagem capazes de produzir texto fluente a partir de instruções em linguagem natural, emergiu, a partir de 2022, como uma das tecnologias de adoção mais acelerada observadas em ambientes organizacionais. Diferentemente de soluções de tecnologia da informação tradicionais, muitas dessas ferramentas são gratuitas, acessíveis por um navegador e dispensam competência técnica especializada, o que permite que qualquer servidor as utilize de forma imediata e individual, à margem das estruturas formais de governança da sua organização.

Quando agentes públicos recorrem a essas ferramentas por iniciativa própria, sem autorização, registro ou supervisão institucional, configura-se o que a literatura recente tem denominado *shadow AI,* extensão, para o domínio da inteligência artificial, do conceito consolidado de *shadow IT*, entendido como o uso de soluções tecnológicas não aprovadas pelo setor de tecnologia da organização (Silic e Back, 2014). Como observam Silic, Silic e Kind-Trüller (2025), o fenômeno caracteriza-se, em regra, não por má-fé, mas como ameaça interna não maliciosa, motivada por pressões de produtividade e pela ausência de alternativas institucionais. A disseminação da prática já encontra respaldo empírico em pesquisas recentes sobre o setor público, que documentam adoção crescente sem correspondência em orientação institucional clara (Bright *et al.*, 2024).

Na tradição burocrática weberiana, o uso informal de inteligência artificial generativa adquire relevância institucional porque a decisão administrativa fundamenta-se no registro documental, memória institucional que assegura consistência, permite a auditoria dos atos e sobrevive ao agente individual (Weber, 2009). Essas ferramentas são suscetíveis à alucinação, ou seja, à geração de conteúdo fluente, porém factualmente incorreto ou sem sustentação em evidência (Huang *et al.*, 2025), introduzindo no processo decisório informações de origem epistêmica opaca. Adicionalmente, ao ocorrer sem registro, a mediação algorítmica permanece invisível, de modo que não se sabe qual instrução gerou determinado conteúdo, qual modelo o produziu ou se houve verificação. Compromete-se, assim, tanto a integridade informacional quanto a possibilidade de responsabilização dos atos.

Diante desse cenário, a questão deste estudo é quais riscos institucionais a literatura associa ao uso informal de inteligência artificial generativa para a integridade informacional e a *accountability* na administração pública. Para respondê-la, o estudo tem por objetivo analisar os riscos institucionais associados ao uso informal dessas tecnologias na administração pública, especialmente quanto à integridade informacional e à *accountability* dos processos decisórios. Para tanto, caracteriza-se o uso informal, distinguindo-o das iniciativas formalmente regulamentadas, identificam-se os principais riscos institucionais para a integridade informacional e a *accountability* e sistematizam-se as diretrizes de governança apontadas na literatura.

A pertinência da investigação baseia-se no descompasso entre a velocidade de adoção da tecnologia e a maturidade dos mecanismos de governança capazes de disciplinã-la. Embora exista literatura consolidada sobre *shadow IT,* sobre *accountability* algorítmica no setor público como (Bovens, 2007; Busuioc, 2021) e sobre a confiabilidade dos registros (Duranti, 1995), permanecem escassos os trabalhos que articulam esses campos no recorte específico do uso informal de

inteligência artificial generativa por agentes públicos. Essa lacuna mostra-se sensível em um domínio no qual as decisões repercutem sobre direitos dos cidadãos e no qual a exigência de motivação, rastreabilidade e prestação de contas dos atos decorre do regime jurídico-administrativo e do próprio Estado de Direito, e não de mera conveniência gerencial.

O estudo parte da premissa de que práticas individuais informais podem produzir riscos institucionais sobre a integridade informacional e a *accountability* dos processos decisórios públicos, propondo uma contribuição analítica e integradora que articula a Teoria da Burocracia weberiana com a tradição arquivística de Duranti (1995) e com a literatura contemporânea de *accountability* algorítmica (Bovens, 2007; Busuioc, 2021; Bracci, 2023), mobilizando esse arcabouço para interpretar o fenômeno emergente do shadow AI no setor público. Sustenta-se que esses riscos atingem o fundamento documental da racionalidade burocrática e que a integridade informacional condiciona materialmente a *accountability*, e não a acompanha como dimensão paralela. As diretrizes de governança discutidas emergem como decorrência analítica desse diagnóstico, não como prescrição normativa autônoma.

Após esta introdução, o trabalho contempla a revisão teórica e conceitual, a metodologia adotada, a análise e discussão dos resultados e as considerações finais.

2. Fundamentação Teórica

A análise dos riscos institucionais decorrentes do uso informal de inteligência artificial generativa na administração pública requer um arcabouço teórico capaz de articular o fundamento da decisão administrativa, a natureza do caso e as exigências contemporâneas de prestação de contas. Para essa articulação, o presente estudo recorre à Teoria da Burocracia de Weber (2009) como matriz teórica, em diálogo com a literatura de accountability algorítmica (Bovens, 2007; Busuioc, 2021), por oferecer categorias úteis à análise dos efeitos da mediação algorítmica sobre a produção documental e a responsabilização administrativa.

Como observa Weber (2009), a administração burocrática moderna fundamenta-se em normas escritas e em procedimentos formalmente estabelecidos, nos quais decisões, atos e comunicações são lavrados por escrito, constituindo um arquivo institucional que assegura previsibilidade, permite a revisão dos atos e sobrevive aos agentes individuais. Esse princípio do registro constitui elemento da legitimidade racional-legal e da possibilidade de responsabilização, de modo que qualquer prática que insira informações no processo decisório à margem do registro institucional compromete o núcleo do tipo ideal weberiano. Para Bovens (2007), *accountability* constitui uma relação entre um ator e um fórum, na qual o ator tem obrigação de explicar e justificar sua conduta, o fórum pode questioná-lo e julgar tal conduta e o ator pode sofrer consequências. Na visão de Busuioc (2021), a aplicação desse modelo ao contexto da decisão mediada por algoritmos depende da existência de mecanismos efetivos de explicação e justificação, sob pena de inviabilizar a responsabilização.

2.1 Uso informal de IA generativa e o conceito de *shadow AI*

A inteligência artificial generativa designa uma classe de tecnologias sustentadas por grandes modelos de linguagem capazes de produzir conteúdo textual a partir de instruções em linguagem natural. Sua adoção em ambientes organizacionais acelerou-se a partir de 2022, com o lançamento de ferramentas de

uso amplo como o *ChatGPT,* e apresenta característica distintiva em relação às tecnologias de informação tradicionais, marcadas pela gratuidade de acesso, baixa exigência técnica para utilização e independência em relação à infraestrutura institucional. Esse perfil de acesso permite que qualquer agente público utilize as ferramentas de forma imediata, individual e à margem das estruturas formais de governança da sua organização.

A caracterização desse uso encontra base teórica na literatura de *shadow IT,* na qual Silic e Back (2014), o conceito designa o emprego de soluções de *hardware* ou *software* por colaboradores sem aprovação formal do setor de tecnologia da organização, fato tratado pelos autores como ameaça interna não maliciosa, motivada por pressões de produtividade e pela ausência de alternativas institucionais. A categorização é relevante porque dissocia o risco da intencionalidade do agente, uma vez que o problema não decorre de má-fé, mas da própria forma de inserção da tecnologia na rotina de trabalho. Em desenvolvimento mais recente, Silic, Silic e Kind-Trüller (2025) propõem a extensão analítica desse conceito para o domínio da inteligência artificial, cunhando a noção de *shadow AI* como o uso não sancionado de sistemas de IA fora dos marcos de governança aprovados pela organização

A continuidade conceitual entre os dois casos sustenta-se na lógica do contorno institucional, mas há um deslocamento qualitativo relevante do *shadow IT* para o *shadow AI* que merece destaque. O *shadow IT* clássico incide essencialmente sobre a infraestrutura tecnológica utilizada pelo colaborador, como o emprego de plataformas de armazenamento, comunicação ou compartilhamento não autorizadas, situadas no entorno operacional da decisão. O *shadow AI,* por sua vez, atinge a produção do próprio conteúdo administrativo, isto é, a inteligência artificial generativa passa a participar da elaboração de pareceres, relatórios e decisões, e essa participação ocorre fora do registro institucional. O ponto de inflexão analítico não está mais na infraestrutura paralela, mas na introdução de uma mediação algorítmica invisível no próprio núcleo do trabalho intelectual da burocracia, dimensão ainda pouco explorada na literatura.

Para os fins desta pesquisa, distingue-se o uso institucional formal de IA generativa no setor público, isto é, soluções deliberadamente adotadas, regulamentadas, contratadas e auditadas pela organização, do uso informal por agentes públicos, caracterizado pela utilização individual de ferramentas externas sem autorização, registro ou supervisão institucional. Pereira e Araújo (2025) ilustram o primeiro caso ao analisar a adoção institucional de IA, inclusive generativa, pelos Tribunais de Contas brasileiros, na qual a tecnologia é incorporada por meio de políticas, capacitações e instrumentos avaliativos formais. O fenômeno aqui estudado situa-se na outra ponta dessa dinâmica, o uso informal, que na visão de Silic, Silic e Kind-Trüller (2025), configura um contorno do princípio burocrático do registro, na medida em que a mediação algorítmica passa a integrar a produção do conteúdo administrativo sem deixar trilha auditável.

2.2 Integridade informacional e *accountability* como dimensões em risco

A análise dos riscos do uso informal de inteligência artificial generativa parte do reconhecimento de que a integridade informacional opera como condição material da *accountability*, pois, na ausência de atributos que assegurem a qualidade e a verificabilidade das informações que fundamentam o ato administrativo, a obrigação formal de explicá-lo e justificá-lo perde sua substância. Portanto, os conceitos não são paralelos, mas hierarquicamente articulados.

Assumida essa relação, entende-se a integridade informacional como a condição de confiabilidade, autenticidade, fidedignidade e rastreabilidade das informações utilizadas na fundamentação de decisões administrativas, de modo que seja possível identificar a origem, o percurso, os critérios de validação e eventuais mediações, inclusive algorítmicas, a que essas informações foram submetidas. Essa compreensão apoia-se na tradição arquivística da diplomática contemporânea, na qual Duranti (1995) conceitua a confiabilidade do registro como a capacidade de uma informação documental servir como fato em razão dos procedimentos adotados em sua criação, e a autenticidade como a condição de o registro ser aquilo que pretende ser e estar livre de adulteração ou corrupção. Esse embasamento distingue o conceito aqui empregado das noções correlatas de integridade da informação em segurança cibernética e de integridade pública em sentido ético-anticorrupção.

A integridade informacional encontra-se em risco no contexto do uso informal de IA generativa por uma combinação de fatores. Inicialmente, pela suscetibilidade dos próprios modelos à alucinação, descrita por Huang *et al.* (2025), como produção de conteúdo fluente e sintaticamente coerente, mas factualmente incorreto ou desprovido de sustentação em evidência externa, ocorrência que compromete a confiabilidade do material gerado, sobretudo em domínios que exigem precisão factual. Quando inserido informalmente no processo decisório público, esse conteúdo passa a operar como informação validada, sem que haja registro do processo que o produziu. Outro fator é o apagamento da mediação algorítmica, pois a ausência de registro do prompt utilizado, do modelo acionado e dos eventuais ajustes humanos posteriores, torna-se impossível reconstruir a cadeia de produção da informação, o que rompe diretamente o componente de rastreabilidade do conceito aqui adotado.

Esses movimentos repercutem na *accountability* dos atos administrativos, na medida em que a relação de *accountability* como aponta Bovens (2007), pressupõe que o ator forneça ao fórum elementos para explicar e justificar sua conduta, condição que se inviabiliza quando parte substantiva da fundamentação do ato decorre de um processo que o próprio ator não pode reconstituir. Na visão de Busuioc (2021), essa dificuldade adquire feição específica no caso da decisão algorítmica, na medida em que a opacidade dos sistemas e a ausência de trilhas auditáveis tornam vazia a obrigação formal de justificação. Em desenvolvimento convergente, Bracci (2023) cunha a noção de brechas estruturais que se abrem na cadeia de prestação de contas quando o setor público incorpora decisões mediadas por algoritmos, propondo agenda específica para o tratamento dessas lacunas. No setor público, a perda da rastreabilidade não constitui apenas falha gerencial, pois compromete o dever de motivação dos atos administrativos, exigência decorrente do regime jurídico-administrativo e do Estado de Direito.

2.3 Governança de IA generativa no setor público

Se a integridade informacional constitui condição material da *accountability*, os mecanismos de governança podem ser compreendidos como tentativas institucionais de recompor as condições de rastreabilidade e responsabilização que sustentam a legitimidade burocrática. Em outras palavras, a governança da IA generativa não se destina apenas ao controle da tecnologia, mas à reconstrução das condições que tornam possível a responsabilização dos atos administrativos.

Sob essa perspectiva, os marcos normativos contemporâneos procuram traduzir em instrumentos operacionais exigências associadas à transparência, à explicabilidade e à responsabilização. O NIST (2024) propõe um perfil de gestão de

risco específico para IA generativa, no qual orienta organizações a mapear, medir e gerenciar riscos associados à produção e ao uso desses sistemas, incluindo a necessidade de documentação dos prompts, dos modelos empregados e das saídas geradas. Em direção convergente, a OECD (2019) estabelece princípios de IA centrados em valores como transparência, confiabilidade técnica, *accountability* e respeito ao Estado de Direito, posteriormente atualizados em 2024 para incorporar especificidades da IA generativa.

No contexto brasileiro, o Ministério da Gestão e da Inovação em Serviços Públicos (Brasil, 2024) publicou cartilha orientativa sobre o uso de IA generativa no serviço público federal, na qual aponta práticas recomendadas, riscos a serem observados e a necessidade de validação humana das saídas algorítmicas antes de sua incorporação a documentos administrativos. Em construção complementar a essa orientação operacional, tramita no Congresso Nacional o Projeto de Lei nº 2.338 (Brasil, 2023).

A literatura empírica evidencia que mecanismos formais e informais de governança coexistem nas organizações (Vallé *et al*., 2026), e Ulnicane (2025) observa, no setor público, um paradoxo na governança da IA generativa, na medida em que se trata de tecnologia em uso público amplo, mas com governança restrita a um número limitado de centros decisórios. Esse paradoxo evidencia que a expansão do uso da IA generativa não foi acompanhada por mecanismos equivalentes de responsabilização. Nesse contexto, a governança deixa de representar mera camada adicional de controle tecnológico e passa a constituir esforço institucional voltado à recomposição das condições de rastreabilidade e prestação de contas que sustentam a racionalidade burocrática.

3. Metodologia

A pesquisa caracteriza-se por uma abordagem qualitativa, de natureza aplicada e objetivo exploratório-descritivo, tendo adotado a revisão integrativa, adequada a temas emergentes e à combinação de evidências empíricas, teóricas e normativas, em síntese qualitativa unificada (Souza; Silva; Carvalho, 2010; Mendes; Silveira; Galvão, 2008; Whittemore; Knafl, 2005).

A busca foi conduzida em maio de 2026 nas bases Scopus, Web of Science e SciELO, combinando descritores referentes à tecnologia, ao setor público e aos fenômenos de *shadow AI, shadow IT*, *accountability* algorítmica e governança de IA. As buscas foram realizadas nos campos de título, resumo e palavras-chave, em português, inglês e espanhol, restritas a publicações de 2022 em diante, marco da popularização da IA generativa, e a artigos revisados por pares, capítulos, anais de eventos e revisões. Considerando que os termos shadow *AI* e uso informal ainda não constituem descritores consolidados, ampliou-se o bloco do fenômeno para captar estudos pertinentes sob denominações correlatas, ancorando-se o conceito na literatura de *shadow IT* (Silic; Back, 2014). Acrescentaram-se ao corpus, por busca complementar dirigida, o levantamento empírico de Bright *et al.* (2024) sobre a difusão da IA generativa no setor público britânico e o estudo de Ribeiro e Segatto (2025) sobre a adoção de inteligência artificial em organizações públicas brasileiras.

A estratégia de busca articulou descritores relativos à tecnologia focal, ao recorte institucional do setor público e aos fenômenos e riscos associados ao uso informal, conforme apresentado no Quadro 1 que sintetiza essa arquitetura.

Quadro 1- Estratégia de busca por bloco de descritores

Bloco	Descritores (combinados por OR)
Tecnologia	*generative AI; generative artificial intelligence; large language model; LLM; ChatGPT; IA generativa*
Setor público	*public sector; public administration; government; civil servant; public official; administração pública; servidor público.*
Fenômeno e riscos	*shadow AI; shadow IT; unsanctioned AI; unauthorized AI; uso informal; algorithmic accountability; AI governance; information integrity.*

Fonte: elaborado pelas autoras (2026).

A combinação de descritores consolidados, como *shadow IT*, com termos ainda em consolidação na literatura, como *shadow AI* e *unsanctioned AI,* respondeu à novidade do tema e ampliou a chance de captura de estudos pertinentes. A inclusão simultânea de termos em português, inglês e espanhol procurou reduzir vieses linguísticos na cobertura das bases.

Como critérios de inclusão, mantiveram-se os estudos que tratavam simultaneamente dos três blocos conceituais, excluindo-se trabalhos sobre IA em sentido amplo sem recorte de setor público, trabalhos sobre setor público sem relação com IA generativa, notícias e opiniões sem fundamentação. A seleção foi realizada em duas etapas sucessivas de triagem, por título e resumo e por leitura integral, com reporte conforme o PRISMA 2020 (Page *et al*., 2021). A busca identificou 64 registros nas bases (27 na Scopus, 36 na Web of Science e 1 na SciELO), aos quais se somaram dois registros da busca complementar. Não foram detectadas duplicatas entre as bases. Após a triagem, compuseram o corpus final 21 estudos.

A extração dos dados foi realizada em matriz de síntese registrada em planilha eletrônica, contendo a identificação, o contexto, a natureza da evidência, as formas de uso descritas, os riscos identificados às dimensões analíticas da pesquisa, as diretrizes de governança propostas e observações pertinentes. O registro da natureza da evidência, classificada como empírica, teórica ou normativa, permitiu distinguir riscos empiricamente observados daqueles apenas projetados teoricamente.

Os dados foram submetidos à análise de conteúdo categorial (Bardin, 2011), com categorias definidas a priori a partir dos eixos analíticos do estudo e categorias emergentes construídas durante a leitura. As categorias prévias contemplaram os componentes do conceito operacional de integridade informacional, as dimensões do modelo ator-fórum de *accountability* (Bovens, 2007) e os tipos de mecanismo de governança identificados nos marcos normativos. As categorias emergentes captaram riscos não plenamente enquadráveis nas categorias prévias, como a exposição de dados sob sigilo legal. As diretrizes extraídas foram confrontadas com os marcos de referência do NIST (2024), da OECD (2019) e do Brasil (2024).

4. Análise e Discussão dos Resultados

O corpus analisado reuniu 21 estudos publicados entre 2024 e 2026, refletindo a fase inicial da produção científica sobre o tema. A distribuição geográfica revela predomínio de pesquisas conduzidas em contextos do hemisfério norte, com investigações em órgãos públicos dos Estados Unidos (Chen; Gasco-Hernandez; Gil-Garcia, 2026; Hatz *et al.*, 2025), na Comissão Europeia (De Longueville *et al.*, 2025), na Austrália (Lozano-Paredes, 2025) e no Reino Unido (Bright *et al.*, 2024), complementadas por pesquisas em contextos de países em desenvolvimento, como

Vietnã (Duy *et al*., 2026), Arábia Saudita (Alshahrani *et al.*, 2025) e Brasil (Pereira; Araújo, 2025).

Quanto à natureza da evidência, 12 estudos apresentaram desenho empírico, com coleta de dados primários por meio de *surveys*, estudos de caso ou análise documental, e os 9 restantes adotaram caráter teórico, conceitual ou propositivo. Esse equilíbrio relativo entre evidência empírica e produção conceitual é relevante para a análise, pois permite distinguir riscos efetivamente constatados em campo daqueles apenas projetados pela literatura.

4.1 Caracterização do uso informal de IA generativa na administração pública

O exame conjunto do corpus permite caracterizar o uso informal de IA generativa como antecedente estrutural do tensionamento na cadeia documental burocrática. Os estudos empíricos analisados indicam que esse uso por agentes públicos já configura prática difundida, cuja incidência supera, com folga, a maturidade dos arranjos institucionais que deveriam disciplinà-la. O levantamento de Bright *et al.* (2024), conduzido com 938 profissionais do setor público britânico, indica que 22% dos respondentes utilizavam ativamente ferramentas de IA generativa em suas atividades de trabalho, enquanto apenas 32% percebiam orientação institucional clara sobre o uso.

Esse descompasso entre adoção real e governança formal aparece de modo recorrente em outras pesquisas empíricas, como em Tangi, Müller e Combetto (2026), que descrevem a presença de sombra da IA generativa em administrações públicas europeias, e em Frisch-Aviram, Spanghero Lotta e Jordão de Carvalho (2025), que documentam empiricamente a disposição de burocratas de nível de rua para quebrar regras tecnológicas formais quando submetidos a pressões competitivas. O padrão sugere tratar-se de fenômeno estrutural, e não anedótico.

As formas concretas pelas quais a tecnologia se insere na rotina administrativa convergem para um conjunto restrito de práticas, em geral relativas à produção textual administrativa. A redação de relatórios, a síntese de documentos, o refino de propostas e a elaboração de pareceres aparecem como atividades em que a IA generativa é mobilizada por iniciativa individual do agente, sem registro institucional, como documentado por Tangi, Müller e Combetto (2026), Neumann *et al.* (2026) e Alshahrani *et al.* (2025). Ao operar precisamente sobre a etapa de produção do conteúdo administrativo, o *shadow AI* tensiona o princípio weberiano segundo o qual decisões e comunicações da burocracia devem ser lavradas por escrito e ingressar no arquivo institucional como base da legitimidade racional-legal (Weber, 2009). A mediação algorítmica passa a integrar a elaboração documental sem deixar nele a trilha do procedimento adotado em sua criação, condição que Duranti (1995) identifica como pressuposto da confiabilidade do registro.

A literatura identifica motivações estruturais, não comportamentais, para a difusão dessa prática, descrita por Neumann *et al.* (2026) em termos de *policy-practice* gaps, isto é, descompassos entre políticas formais e padrões reais de uso, agravados pela ausência de alternativas institucionais validadas. O diagnóstico converge com a noção de ameaça interna não maliciosa consolidada na literatura de *shadow IT* (Silic; Back, 2014) e ressaltada por Silic, Silic e Kind-Trüller (2025) na sua extensão para o *shadow AI.* O agente público não é ator desviante, mas funcionário pressionado por demandas de produtividade em ambiente organizacional que ainda não ofereceu alternativa formal ao recurso que ele encontra disponível no navegador.

O corpus oferece também o contraponto, ao reunir casos em que a inteligência artificial generativa é incorporada de modo institucional formal. Pereira e Araújo (2025) descrevem a adoção da tecnologia pelos Tribunais de Contas brasileiros, mediada por políticas, capacitações e instrumento avaliativo de maturidade institucional, e Papageorgiou *et al.* (2024) detalham uma arquitetura modular baseada em *large language model* - LLM para serviços de governo eletrônico na Grécia, com geração aumentada por recuperação tecnicamente auditável.

Ribeiro e Segatto (2025) acrescentam, com base nos microdados da Pesquisa TIC Governo Eletrônico 2021, que essa adoção institucional permanece restrita a um grupo reduzido de entidades brasileiras e fortemente heterogênea entre entes federais e estaduais, padrão que ajuda a explicar a propensão ao recurso informal por parte de servidores em organizações sem cobertura institucional adequada. A experiência *GPT@JRC,* na Comissão Europeia, descrita por De Longueville *et al.* (2025), indica que o reconhecimento precoce da difusão informal permite seu deslocamento para canais auditáveis, o que sugere caminho possível, ainda pouco trilhado no plano nacional.

4.2 Riscos institucionais à integridade informacional e à *accountability*

Os riscos institucionais que emergem do corpus organizam-se em torno da dependência da accountability em relação à integridade informacional. Os achados sustentam essa dependência ao mostrar que o comprometimento da integridade informacional inviabiliza, em substância, a prestação de contas, ainda que suas formas institucionais permaneçam nominalmente preservadas. O Quadro 2 organiza os riscos institucionais identificados na revisão em dois eixos analíticos, distinguindo o mecanismo causal subjacente, a natureza da evidência que o sustenta e os principais estudos do corpus que dão suporte ao achado.

.

Quadro 2 - Síntese dos riscos institucionais associados ao uso informal de IA generativa no setor público

Eixo	Risco identificado	Mecanismo causal	Estudos principais do corpus/Natureza (Empírica= [E] ; Teórica = [T]
Integridade informacional	Comprometimento da fidedignidade	Alucinação dos modelos	Ahmad *et al.* (2026) [T]; Tangi, Müller e Combetto (2026) [E]; Hinov e Ivanova (2026) [T].
Integridade informacional	Apagamento da rastreabilidade	Ausência de registro do prompt, modelo acionado e ajustes humanos posteriores	Hinov e Ivanova (2026) [T]; Tangi, Müller e Combetto (2026) [E]; Hatz *et al.* (2025) [E].
Integridade informacional	Exposição de dados sob sigilo	Inserção de informação protegida em sistemas comerciais externos	Neumann *et al.* (2026) [E]; Alshahrani *et al*. (2025) [E]; De Longueville *et al.* (2025) [E].
Accountability	Indefinição operacional do modelo ator-fórum	Ausência de definição clara sobre quem responde, perante qual fórum e por quais padrões	Chen, Gasco-Hernandez e Gil-Garcia (2026) [E]; Hatz *et al.* (2025) [E].
Accountability	Paradoxo da governança	Uso público amplo coexistindo com governança institucional restrita a poucos centros decisórios	Ulnicane (2025) [T]; Lozano-Paredes (2025) [E]; Duy *et al.* (2026) [E].

Fonte: Elaborado pelas autoras com base na revisão integrativa.

A distribuição dos achados indica que os riscos à integridade informacional apoiam-se de modo predominante em evidência empírica, enquanto os riscos à accountability combinam achados empíricos e diagnóstico teórico, padrão que reflete a fase inicial da produção sobre o tema.

No eixo da integridade informacional, os achados corroboram que a alucinação dos modelos descrita por Huang *et al.* (2025) opera como mecanismo causal central da fragilização da fidedignidade, cujas implicações em contextos institucionais Ahmad *et al.* (2026) demonstram no campo da auditoria. Quando o conteúdo gerado é inserido informalmente no fluxo administrativo, passa a operar como informação validada, ainda que o procedimento que o produziu seja inacessível. Essa opacidade rompe diretamente o componente de rastreabilidade da integridade informacional e contraria o princípio de Duranti (1995), segundo o qual a confiabilidade do registro deriva dos procedimentos adotados em sua criação. Acrescenta-se a esse quadro a exposição de dados sob sigilo legal a sistemas comerciais externos, registrada em estudos como Neumann *et al.* (2026), Alshahrani *et al.* (2025) e De Longueville *et al.* (2025), risco cuja magnitude no contexto público ainda carece de mensuração empírica sistemática, embora seu efeito teórico, sobre o princípio weberiano do arquivo institucional como elemento que sobrevive ao agente individual (Weber, 2009), seja inequívoco.

No eixo da *accountability*, a indefinição operacional do modelo ator-fórum constitui achado central da análise, como verificado por Chen, Gasco-Hernandez e Gil-Garcia (2026) a maior parte dos instrumentos institucionais examinados não define com clareza quem é responsabilizado, perante quem, por quais padrões e com que consequências, achado coerente com o diagnóstico teórico de Bracci (2023) sobre as brechas estruturais que se abrem na incorporação de decisões algorítmicas e com a posição de Busuioc (2021) sobre o esvaziamento das obrigações formais de justificação diante de sistemas opacos. Hatz *et al.* (2025) corroboram o quadro ao demonstrar que, embora gestores locais reconheçam a importância da prestação de contas no uso de IA, os mecanismos concretos permanecem subdesenvolvidos. Ulnicane (2025) inscreve esses resultados em camada macropolítica ao retomar o paradoxo da governança discutido na Seção 2.3, padrão também observado em escala local por Lozano-Paredes (2025) na Austrália e Duy *et al.* (2026) no Vietnã.

A leitura conjunta confirma que o comprometimento da integridade informacional produzido pelo *shadow AI* atinge o fundamento documental da racionalidade burocrática weberiana, ultrapassando a esfera da falha gerencial pontual. Ao mediar invisivelmente a produção do conhecimento que fundamenta a decisão, o uso informal esvazia a substância da prestação de contas.

4.3 Diretrizes de governança para mitigação dos riscos

Os mecanismos de governança identificados na literatura não aparecem como prescrições independentes dos riscos anteriormente descritos. Ao contrário, emergem como tentativas institucionais de recompor os elementos da integridade informacional comprometidos pelo uso informal da IA generativa e, por consequência, restaurar as condições materiais da *accountability*.

Em resposta à opacidade da rastreabilidade, os marcos normativos e a literatura analítica convergem para a recomendação de documentação sistemática da interação entre agente e modelo. O NIST (2024), em seu perfil de gestão de risco para IA generativa, prescreve o registro de prompts, modelos acionados e saídas geradas

como condição mínima de governabilidade do risco. Hinov e Ivanova (2026) sustentam, em paralelo, que arquiteturas institucionais devem incorporar logs estruturados das interações que viabilizem reconstrução posterior do processo decisório, e Ahmad *et al.* (2026) reforçam essa diretriz ao demonstrar que trilhas de auditoria são condição necessária à preservação da evidência documental em contextos de uso de IA generativa. Mais do que exigências técnicas, esses mecanismos representam tentativas de restabelecer a cadeia documental cuja ruptura foi identificada como principal consequência institucional do *shadow AI.*

Em relação à indefinição dos polos da relação de *accountability*, a OECD (2019) consolida princípios de transparência, confiabilidade técnica e responsabilização aplicáveis ao desenho institucional, complementados por Ganapati e Desouza (2024), que defendem a governança de IA no setor público ancorada em princípios de valor público com ênfase na contestabilidade dos componentes algorítmicos. Alyileili e Opoku (2026) acrescentam que a transparência sobre a mediação algorítmica em serviços públicos digitais deve ser oferecida ao cidadão como elemento estruturante.

No Brasil, a cartilha do Ministério da Gestão e da Inovação em Serviços Públicos (Brasil, 2024) consolida essas orientações com ênfase na validação humana das saídas algorítmicas antes de sua incorporação a documentos administrativos. Esses instrumentos procuram responder à necessidade de tornar identificáveis os pólos da relação ator-fórum, problema identificado por Bovens (2007). A transparência sobre a mediação algorítmica e a preservação da supervisão humana constituem mecanismos destinados a manter a possibilidade de explicação e justificação dos atos administrativos. Contudo, os diagnósticos de Ulnicane (2025) e Jonathan e Han (2026) indicam que as recomendações atualmente disponíveis permanecem genéricas e insuficientemente operacionalizadas, evidenciando que a governança da IA generativa ainda se encontra em fase de consolidação, marcada por soluções mais principiológicas do que operacionalmente maduras.

5. Conclusão

A literatura analisada associa ao uso informal de IA generativa por agentes públicos um conjunto de riscos institucionais que afeta tanto a integridade informacional quanto a *accountability* dos processos decisórios. A tese sustentada a partir do exame conjunto desses achados é que o *shadow AI* não é apenas problema tecnológico ou gerencial, mas institucional: ao introduzir uma mediação algorítmica que não ingressa no arquivo institucional, rompe a cadeia documental da racionalidade burocrática e esvazia a obrigação de explicar e justificar o ato administrativo.

A contribuição do trabalho assenta-se na articulação entre Weber, Duranti e a literatura contemporânea de *accountability* algorítmica, aplicada ao fenômeno emergente do *shadow AI* no setor público. Essa articulação demonstra que a integridade informacional é condição material para *accountability*, ou seja, sem informação confiável e rastreável na fundamentação do ato, a obrigação de prestar contas opera no vazio. O estudo evidencia ainda que o *shadow AI* representa deslocamento qualitativo em relação ao *shadow IT* clássico, pois a inserção informal não se limita à infraestrutura, mas penetra a produção do próprio conhecimento administrativo. As diretrizes de governança emergem como decorrência analítica desse diagnóstico, não como cartilha normativa autônoma.

O estudo apresenta limitações como o tamanho do corpus que reflete o caráter emergente do tema e a inexistência, nas plataformas consultadas, de literatura

indexada que articulasse simultaneamente uso informal, inteligência artificial generativa e administração pública. A predominância de pesquisas conduzidas em países desenvolvidos restringe a transferência direta dos achados ao contexto brasileiro, no qual a base de evidências indexadas permanece escassa. A revisão integrativa, ainda que adequada à imaturidade do campo, não substitui a investigação empírica primária necessária para captar dimensões específicas do uso informal em organizações públicas nacionais.

Sugere-se, para pesquisas futuras, a condução de estudos empíricos qualitativos em órgãos da administração pública brasileira, com o propósito de mapear as formas concretas pelas quais a IA generativa tem sido incorporada por agentes públicos à margem das estruturas institucionais. Recomenda-se também o desenvolvimento de articulação específica entre os riscos identificados e o dever constitucional de motivação dos atos administrativos, dimensão pouco explorada na literatura internacional analisada. No plano regulatório nacional, a tramitação do Projeto de Lei nº 2.338/2023, em apreciação na Câmara dos Deputados em 2026, sinaliza a construção de marco regulatório brasileiro de inteligência artificial, contexto no qual a discussão sobre uso informal e seus riscos institucionais ganha pertinência regulatória imediata (Brasil, 2023).

Ao tornar invisível a mediação algorítmica na produção do conhecimento administrativo, o *shadow AI* recoloca uma questão clássica da burocracia, que é a preservação do registro documental como condição da legitimidade e da responsabilização dos atos do Estado. A sistematização apresentada contribui para o debate acadêmico, para a formulação de respostas institucionais por parte de gestores públicos e instâncias de controle e para a maturação de marcos regulatórios sensíveis às especificidades do regime jurídico-administrativo brasileiro.

REFERÊNCIAS

AHMAD, S.; HARBOLA, A.; SACHDEVA, S.; YADAV, B. K.; SHARMA, M. K.; KHAN, A.; AHMAD, J. Generative AI in enterprise IT: implications for audit evidence, assurance, and governance. **EDPACS**, 2026. DOI: https://doi.org/10.1080/07366981.2026.2633846. Acesso em: 10 maio 2026.

ALSHAHRANI, Albandari; GRIVA, Anastasia; DENNEHY, Denis; MÄNTYMÄKI, Matti. The role of leadership and communication in AI assimilation: case studies from Saudi Arabia's public sector organizations. **Transforming Government: People, Process and Policy**, 2025. DOI: https://doi.org/10.1108/TG-06-2025-0187. Acesso em: 10 maio 2026.

ALYILEILI, M.; OPOKU, A. Sustainable and inclusive AI governance in municipal self-service systems: ethical, smart-government, and generative AI perspectives. **Sustainability**, v. 18, n. 2, art. 849, 2026. DOI: https://doi.org/10.3390/su18020849. Acesso em: 8 maio 2026.

BARDIN, Laurence. **Análise de conteúdo**. Tradução de Luís Antero Reto e Augusto Pinheiro. São Paulo: Edições 70, 2011.

BOVENS, Mark. Analysing and assessing accountability: a conceptual framework. **European Law Journal**, v. 13, n. 4, p. 447-468, 2007. DOI: https://doi.org/10.1111/j.1468-0386.2007.00378.x. Acesso em: 6 maio 2026.

BRACCI, Enrico. The loopholes of algorithmic public services: an "intelligent" accountability research agenda. **Accounting, Auditing & Accountability Journal**, v. 36, n. 2, p. 739-763, 2023. DOI: https://doi.org/10.1108/AAAJ-06-2022-5856. Acesso em: 6 maio 2026.

BRASIL. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 2 jun. 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **IA generativa no serviço público**: definições, usos e boas práticas. Brasília: MGI, 2024. Disponível em: https://www.gov.br/governodigital. Acesso em: 12 jun. 2026.

BRIGHT, Jonathan; ENOCK, Florence E.; ESNAASHARI, Saba; FRANCIS, John; HASHEM, Youmna; MORGAN, Deborah. **Generative AI is already widespread in the public sector**. London: Alan Turing Institute, 2024. arXiv:2401.01291. Disponível em: https://arxiv.org/abs/2401.01291. Acesso em: 12 jun. 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. Acesso em: 4 maio 2026.

CHEN, T. H.; GASCO-HERNANDEZ, M.; GIL-GARCIA, J. R. Understanding accountability in AI use within the public sector: insights from state governments in the United States. **Public Administration**, 2026. DOI: https://doi.org/10.1111/padm.70051. Acesso em: 4 maio 2026.

DE LONGUEVILLE, B.; SANCHEZ, I.; KAZAKOVA, S.; LUONI, S.; ZARO, F.; DASKALAKI, K.; INCHINGOLO, M. The proof is in the eating: lessons learnt from one year of generative AI adoption in a science-for-policy organisation. **AI**, v. 6, n. 6, art. 128, 2025. DOI: https://doi.org/10.3390/ai6060128. Acesso em: 7 maio 2026.

DURANTI, Luciana. Reliability and authenticity: the concepts and their implications. **Archivaria**, n. 39, p. 5-10, 1995. Disponível em: https://archivaria.ca/index.php/archivaria/article/view/12063. Acesso em: 2 maio 2026.

DUY, P. N.; RUANGTHAMSING, C.; KAMNUANSILPA, P.; LOWATCHARIN, G.; SETTHASURAVICH, P. Generative AI in developing countries: adoption dynamics in Vietnamese local government. **Informatics**, v. 13, n. 2, art. 22, 2026. DOI: https://doi.org/10.3390/informatics13020022. Acesso em: 4 maio 2026.

FRISCH-AVIRAM, N.; SPANGHERO LOTTA, G.; JORDÃO DE CARVALHO, L. "Chat-Up": the role of competition in street-level bureaucrats' willingness to break technological rules and use generative pre-trained transformers (GPTs). **Public Administration Review**, 2025. DOI: https://doi.org/10.1111/puar.13824. Acesso em: 8 maio 2026.

GANAPATI, Sukumar; DESOUZA, Kevin C. Public value principles for secure and trusted AI. In: ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH, 25., 2024, Taipei. **Anais** [...]. New York: ACM, 2024. DOI: https://doi.org/10.1145/3657054.3657086. Acesso em: 8 maio 2026.

HATZ, Sophia; DREKSLER, Noemi; WEI, Kevin; ZHANG, Baobao. Local US officials' views on the impacts and governance of AI: evidence from 2022 and 2023 survey waves. **PLoS ONE**, v. 20, n. 10, art. e0332919, 2025. DOI: https://doi.org/10.1371/journal.pone.0332919. Acesso em: 8 maio 2026.

HINOV, N.; IVANOVA, M. LLM-augmented algorithmic management: a governance-oriented architecture for explainable organizational decision systems. **AI**, v. 7, n. 3, art. 102, 2026. DOI: https://doi.org/10.3390/ai7030102. Acesso em: 8 maio 2026.

HUANG, Lei; YU, Weijiang; MA, Weitao; ZHONG, Weihong; FENG, Zhangyin; WANG, Haotian; CHEN, Qianglong; PENG, Weihua; FENG, Xiaocheng; QIN, Bing; LIU, Ting. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. **ACM Transactions on Information Systems**, v. 43, n. 2, art. 42, 2025. DOI: https://doi.org/10.1145/3703155. Acesso em: 8 maio 2026.

JONATHAN, G. M.; HAN, S. N. GenAI in public sector transformation: balancing promise and prudence. **Lecture Notes in Information Systems and Organisation**. Cham: Springer, 2026. p. 95-108. DOI: https://doi.org/10.1007/978-3-032-02225-7_8. Acesso em: 8 maio 2026.

LOZANO-PAREDES, L. Mapping AI's role in NSW governance: a socio-technical analysis of GenAI integration. **Frontiers in Political Science**, v. 7, art. 1595345, 2025. DOI: https://doi.org/10.3389/fpos.2025.1595345. Acesso em: 8 maio 2026.

MENDES, Karina Dal Sasso; SILVEIRA, Renata Cristina de Campos Pereira; GALVÃO, Cristina Maria. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. **Texto & Contexto Enfermagem**, v. 17, n. 4, p. 758-764, 2008. DOI: https://doi.org/10.1590/S0104-07072008000400018. Acesso em: 8 maio 2026.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **Artificial Intelligence Risk Management Framework**: generative artificial intelligence profile (NIST-AI-600-1). Gaithersburg: NIST, 2024. Disponível em: https://www.nist.gov/itl/ai-risk-management-framework. Acesso em: 12 jun. 2026.

NEUMANN, M.; BISCHOF, L.; HINZ, N. E.; ALTUN, A.; STOCKMANN, L.; SCHRADER, D.; AHAUS, A. C.; DEMIRCI, E. C.; GABEL, B.; RAUSCHENBERGER, M.; DIEBOLD, P.; FRITZEMEIER, H. Between policy and practice: GenAI adoption in agile software development teams. In: GOLDMAN, Alfredo (Ed.). **XP 2026**: Lecture Notes in Business Information Processing, v. 578. Cham: Springer, 2026. p. 290-308. DOI: https://doi.org/10.1007/978-3-032-22375-3_18. Acesso em: 8 maio 2026.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Recommendation of the Council on Artificial Intelligence**. Paris: OECD, 2019 (atualizada em 2024). Disponível em: https://oecd.ai. Acesso em: 12 jun. 2026.

PAGE, Matthew J. *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, v. 372, art. n71, 2021. DOI: https://doi.org/10.1136/bmj.n71. Acesso em: 12 jun. 2026.

PAPAGEORGIOU, G.; SARLIS, V.; MARAGOUDAKIS, M.; TJORTJIS, C. Enhancing e-government services through state-of-the-art, modular, and reproducible

architecture over large language models. **Applied Sciences**, v. 14, n. 18, art. 8259, 2024. DOI: https://doi.org/10.3390/app14188259. Acesso em: 8 maio 2026.

PEREIRA, Flávia Holz Meirelles; ARAÚJO, Luciano Vieira de. O engajamento inovador do controle externo com inteligência artificial: modelo de avaliação. **Revista de Administração Pública**, v. 59, n. 5, e2024-0416, 2025. DOI: https://doi.org/10.1590/0034-761220240416. Acesso em: 8 maio 2026.

RIBEIRO, Manuella Maia; SEGATTO, Catarina Ianni. Inteligência artificial nas organizações públicas brasileiras: heterogeneidades e capacidades em tecnologia da informação. **Revista de Administração Pública**, v. 59, n. 1, 2025. DOI: https://doi.org/10.1590/0034-761220240066. Acesso em: 12 jun. 2026.

SILIC, Mario; BACK, Andrea. Shadow IT: a view from behind the curtain. **Computers & Security**, v. 45, p. 274-283, 2014. DOI: https://doi.org/10.1016/j.cose.2014.06.007. Acesso em: 8 maio 2026.

SILIC, Mario; SILIC, Dario; KIND-TRÜLLER, K. From shadow IT to shadow AI: threats, risks and opportunities for organizations. **Strategic Change**, p. 1-16, 2025. DOI: https://doi.org/10.1002/jsc.2682. Acesso em: 8 maio 2026.

SOUZA, Marcela Tavares de; SILVA, Michelly Dias da; CARVALHO, Rachel de. Revisão integrativa: o que é e como fazer. **Einstein** (São Paulo), v. 8, n. 1, p. 102-106, 2010. DOI: https://doi.org/10.1590/S1679-45082010RW1134. Acesso em: 9 maio 2026.

TANGI, L.; MÜLLER, A. P. R.; COMBETTO, M. A silent partner: the shadow presence of generative artificial intelligence in public administrations. **Lecture Notes in Computer Science**, v. 15978. Cham: Springer, 2026. DOI: https://doi.org/10.1007/978-3-032-02515-9_5. Acesso em: 8 maio 2026.

ULNICANE, Inga. Governance fix? Power and politics in controversies about governing generative AI. **Policy and Society**, v. 44, n. 1, p. 70-84, 2025. DOI: https://doi.org/10.1093/polsoc/puae022. Acesso em: 8 maio 2026.

VALLÉ, R.; SAUP, T.-O.; KANBACH, D. K.; KRAUS, S. The role of formal and informal mechanisms in AI-driven new product development. **Management Decision**, 2026. DOI: https://doi.org/10.1108/MD-10-2025-3324. Acesso em: 7 maio 2026.

WEBER, Max. **Economia e sociedade**: fundamentos da sociologia compreensiva. Tradução de Regis Barbosa e Karen Elsabe Barbosa. Brasília: Editora UnB, 2009. v. 2.

WHITTEMORE, Robin; KNAFL, Kathleen. The integrative review: updated methodology. **Journal of Advanced Nursing**, v. 52, n. 5, p. 546-553, 2005. DOI: https://doi.org/10.1111/j.1365-2648.2005.03621.x. Acesso em: 7 maio 202