Riscos cibernéticos de uso de Inteligência Artificial Generativa no Processo Judicial

Luciana Muniz Costa, Marcus Aurélio Carvalho Georg, Virgínia de Melo Dantas Trinks, Jorilson da Silva Rodrigues, Rafael Rabelo Nunes

Inovações, inteligência artificial e tecnologias de informação e comunicação em Sistemas de Justiça

RESUMO

O Poder Judiciário brasileiro tem incorporado o uso de inteligência artificial em diferentes etapas de sua atuação. A chegada da inteligência artificial generativa (IA-G) amplia tanto as possibilidades quanto os riscos, sobretudo quando aplicada à elaboração de ementas, despachos e decisões judiciais. No contexto jurídico, o emprego dessa tecnologia em tarefas como análise de casos, previsão de resultados e redação de documentos demanda configuração e monitoramento rigorosos, sob pena de introduzir vulnerabilidades relevantes ao processo judicial. O objetivo deste estudo é propor uma metodologia para avaliação de riscos associados ao uso da IA-G na elaboração de despachos e decisões judiciais. Para isso, adotaram-se duas referências principais: um estudo que identifica riscos de negócio relacionados a essa etapa do processo e um framework que sistematiza riscos de IA-G. A metodologia consistiu em aplicar a técnica *bow-tie* para correlacionar os riscos de negócio com os riscos de IA Generativas, permitindo estabelecer conexões entre vulnerabilidades e consequências potenciais. Os resultados indicaram que essa correlação possibilita a identificação de pontos críticos de exposição, bem como a priorização de controles específicos a serem implementados. Como contribuição, o estudo apresenta uma metodologia que pode servir de referência para gestores do Judiciário na seleção de controles prioritários, promovendo maior segurança na adoção da IA-G em atividades sensíveis do processo judicial brasileiro, além de poder ser replicado em outros contextos além do Judiciário.

Palavras-Chave: inteligência artificial generativa; judiciário; riscos cibernéticos; riscos de inteligência artificial generativa.

ABSTRACT

The Brazilian Judiciary has incorporated artificial intelligence into different stages of its operations. The emergence of generative artificial intelligence expands both opportunities and risks, especially when applied to drafting case summaries, judicial orders, and decisions. In the legal context, using this technology for tasks such as case analysis, outcome prediction, and document writing requires strict configuration and monitoring, as it may otherwise introduce significant vulnerabilities into judicial processes. This study aims to propose a methodology for assessing the risks associated with the use of generative artificial intelligence in drafting judicial orders and decisions. To this end, two primary references were adopted: one study that identifies business risks related to this stage of the process and a framework that systematizes generative AI risks. The methodology consisted of applying the bow-tie technique to correlate business risks with generative AI risks, thus establishing connections between vulnerabilities and potential consequences. The results indicated that this correlation enables the identification of critical points of exposure and the prioritization of specific controls to be implemented. As a contribution, the study presents a methodology that can serve as a reference for Judiciary managers in selecting priority controls, fostering greater security

in the adoption of generative artificial intelligence in sensitive judicial activities, and that can also be replicated in contexts beyond the Judiciary.

Keywords: *Generative Artificial Intelligence; Judiciary. Cyber risk; Generative Artificial Intelligence.*

Introdução

Nas últimas décadas, o Poder Judiciário brasileiro tem passado por intensas transformações impulsionadas pela digitalização e pelo uso de tecnologias emergentes. A informatização dos processos judiciais eliminou gradativamente o uso de papel, acelerou a tramitação e modificou a rotina de magistrados, advogados e servidores, promovendo maior eficiência, agilidade e acessibilidade na prestação jurisdicional (Hino & Cunha, 2020).

Nesse contexto, o Brasil tem consolidado um compromisso progressivo com a incorporação de soluções inovadoras, especialmente a inteligência artificial (IA), por meio de políticas e programas que visam modernizar o sistema de justiça (STF, 2023). O Conselho Nacional de Justiça (CNJ) desempenha papel central nesse processo ao cumprir sua competência de planejador estratégico da justiça nacional, por meio da normatização e fomento da aplicação dessas tecnologias, buscando garantir não apenas ganhos de produtividade, mas também transparência, legitimidade e confiança no processo judicial. A recente Resolução CNJ nº 615/2025 estabelece diretrizes para o desenvolvimento, governança e uso responsável de IA no Poder Judiciário, incluindo exigências de auditoria, rastreabilidade e supervisão humana (Conselho Nacional de Justiça, 2025).

A inserção da IA-G no âmbito do poder judiciário nacional amplia as possibilidades de uso, ao permitir a automação de tarefas complexas, como análise de casos, previsão de resultados e elaboração de documentos. Entretanto, sua adoção também potencializa riscos já conhecidos, como os de segurança cibernética, privacidade de dados e vieses algorítmicos, além de introduzir novas vulnerabilidades ainda pouco compreendidas (Alves, Georg & Nunes, 2023). O Perfil de IA-G do National Institute of Standards and Technology (NIST) sistematiza doze riscos específicos dessa tecnologia, incluindo confabulação, recomendações perigosas, toxicidade, viés, homogeneização e uso inadequado de dados pessoais (NIST, 2024).

Experiências recentes ilustram tanto o potencial quanto os desafios dessa inovação. No Supremo Tribunal Federal (STF), o Projeto VICTOR já vinha utilizando IA para classificação de temas de repercussão geral (STF, 2023). Em 2024, o Tribunal de Justiça do Paraná implementou o JurisprudênciaGPT, baseado em IA-G, para automatizar a pesquisa jurisprudencial, iniciativa reconhecida internacionalmente pelo prêmio Gartner Eye on Innovation (CNJ, 2024). Tais avanços demonstram os ganhos de eficiência, mas também ressaltam a necessidade de mecanismos de controle, dado que sistemas generativos estão expostos a ameaças como *prompt injection* e envenenamento de dados, capazes de induzir saídas maliciosas ou enviesadas (Greshake et al., 2023; Wallace et al., 2021). Além disso, a literatura internacional alerta para os riscos de perda de explicabilidade, *deepfakes* e desinformação, que podem comprometer diretamente a legitimidade institucional e a confiança social nas decisões judiciais (Chesney & Citron, 2019; Arrieta et al., 2020).

Embora a literatura sobre riscos de IA tenha avançado nos últimos anos, ainda há escassez de estudos voltados especificamente para o contexto judicial brasileiro, em especial para a etapa de elaboração de ementas, despachos e decisões. Essa lacuna é crítica, pois tais documentos sustentam a legitimidade das decisões judiciais e a própria confiança pública no sistema de justiça. Além disso, a complexidade dos modelos generativos exige metodologias que integrem perspectivas de riscos técnicos e de negócio, de modo a apoiar gestores públicos na definição de controles eficazes e viáveis.

Diante desse cenário, este estudo tem como objetivo propor uma metodologia para avaliação de riscos associados ao uso da IA-G na elaboração de despachos e decisões judiciais. Para isso, foram utilizadas duas referências principais: (i) um estudo que identifica riscos de negócio relacionados a essa etapa do processo e (ii) o framework do NIST que sistematiza riscos de IA-G. A metodologia consistiu em aplicar a técnica *bow-tie* para correlacionar os riscos de negócio com os riscos de IA-G, estabelecendo conexões entre vulnerabilidades e consequências potenciais.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta os principais conceitos relacionados ao estudo; a Seção 3 descreve a metodologia adotada; a Seção 4 correlaciona os riscos de IA-G aos principais riscos de negócio do processo de despachos e decisões no sistema judiciário brasileiro; e, por fim, a Seção 5 apresenta as conclusões e recomendações de trabalhos futuros.

Referencial Teórico

O entendimento conceitual de risco, inteligência artificial e governança é fundamental para sustentar a análise proposta neste estudo. Ao se tratar da aplicação da IA-G no Poder Judiciário, torna-se necessário não apenas definir tais conceitos, mas também explorar suas implicações no campo da segurança cibernética e da legitimidade institucional.

O conceito de risco evoluiu significativamente nas últimas décadas. A norma ABNT NBR ISO 31000:2018 o define como o "efeito da incerteza sobre os objetivos", considerando tanto impactos positivos quanto negativos (ABNT, 2018). Essa abordagem moderna rompe com a visão tradicional de risco como sinônimo exclusivo de ameaça, passando a incluir também oportunidades que emergem de contextos incertos. A gestão de riscos, por sua vez, é entendida como o conjunto de atividades coordenadas para dirigir e controlar uma organização frente a riscos, com o objetivo central de criar e proteger valor. Nesse sentido, a mentalidade de risco permite que organizações antecipem desvios, implementem controles preventivos e maximizem oportunidades. No campo da inteligência artificial, a ABNT NBR ISO/IEC 23894:2023 destaca que os sistemas de IA introduzem riscos emergentes, capazes de alterar a probabilidade e o impacto de riscos já existentes. A norma ABNT ISO/TS 31050:2025 amplia essa visão ao abordar explicitamente a gestão de riscos emergentes para fortalecer a resiliência organizacional diante de tecnologias em rápida evolução.

A inteligência artificial é tradicionalmente definida como o campo da ciência da computação voltado à criação de sistemas que simulam comportamentos humanos, incluindo aprendizado, tomada de decisão e resolução de problemas (Russell & Norvig, 2010). Mais recentemente, a ABNT NBR ISO/IEC 22989:2023 consolidou terminologias que descrevem a IA como disciplina multidisciplinar que envolve ciência da computação, estatística, matemática e até ciências sociais, dada sua interação com valores humanos e

contextos institucionais. Além disso, a ABNT NBR ISO/IEC 42001:2024 propõe um sistema de gestão para IA, ressaltando a importância de diretrizes organizacionais voltadas à segurança, confiabilidade e responsabilidade social. Essa visão amplia a compreensão de IA para além do aspecto técnico, destacando sua necessidade de alinhamento a práticas de governança.

Entre as diferentes categorias de IA, a IA-G tem ganhado destaque por sua capacidade de criar conteúdos, como textos, imagens, áudios e vídeos, a partir de dados de treinamento (NIST, 2024). Diferentemente dos modelos discriminativos, que se limitam a classificar ou prever resultados com base em dados de entrada, os modelos generativos produzem instâncias sintéticas que preservam padrões estatísticos dos dados originais (NIST, 2023). Essa característica confere à IA-G potencial de automação, permitindo desde tarefas rotineiras até a elaboração de minutas complexas em contextos jurídicos. Entretanto, a acessibilidade crescente dessa tecnologia, que possibilita sua utilização até por usuários sem conhecimento técnico avançado, também aumenta o risco de uso inadequado. O Gartner (2024) alerta que a democratização da IA-G pode gerar efeitos adversos, como alucinações factuais, reprodução de vieses e disseminação de desinformação, os quais podem ser reduzidos a partir da implementação de políticas de governança.

A governança de IA pode ser entendida como o conjunto de princípios, práticas e mecanismos voltados a garantir que o desenvolvimento e o uso da tecnologia estejam alinhados a valores éticos, objetivos institucionais e marcos regulatórios. No setor privado, o Instituto Brasileiro de Governança Corporativa (IBGC, 2020) define princípios como integridade, transparência, equidade e responsabilidade corporativa. No setor público, o Decreto nº 9.203/2017 estabelece princípios de governança pública, incluindo capacidade de resposta, integridade, confiabilidade e prestação de contas. No âmbito específico do Poder Judiciário, a Resolução CNJ nº 615/2025 reforça tais princípios ao estabelecer requisitos de supervisão humana, rastreabilidade e segurança para o uso de IA, buscando assegurar legitimidade e confiança social (CNJ, 2025). De forma complementar, o NIST AI RMF (2023) defende que a governança da IA deve abranger todo o ciclo de vida dos sistemas, desde o desenvolvimento até a operação, garantindo confiabilidade, explicabilidade e mitigação de riscos.

A adoção da IA-G apresenta riscos que extrapolam os já conhecidos no campo da tecnologia da informação. O documento NIST AI-600-1 (2024) identifica doze riscos específicos ou agravados por essa tecnologia: confabulação, recomendações perigosas, privacidade de dados, toxicidade, viés e homogeneização, impacto ambiental, integridade da informação, segurança da informação, propriedade intelectual, conteúdo obsceno ou abusivo, configuração humano-IA e vulnerabilidades da cadeia de valor, Tabela 1.

Tabela 1 – Riscos de IA-G

#	Descrição
1	Informações CBRN
2	Confabulação
3	Recomendações Perigosas ou Violentas
4	Privacidade de Dados
5	Impacto Ambiental

6	Configuração Humano-IA
7	Integridade da Informação
8	Segurança da Informação
9	Propriedade Intelectual
10	Conteúdo Obsceno, Degradante e/ou Abusivo
11	Toxicidade, Viés e Homogeneização
12	Cadeia de Valor e Integração de Componentes

Fonte: NIST.AI.600-1 (2024)

Esses riscos não se limitam a questões técnicas, mas alcançam dimensões éticas, sociais e institucionais. No contexto judicial, a presença de informações incorretas, enviesadas ou manipuladas em decisões judiciais pode comprometer diretamente a legitimidade das instituições e a confiança pública. Assim, compreender e sistematizar esses riscos constitui um passo essencial para orientar o desenvolvimento de metodologias de análise e mitigação voltadas ao setor.

Trabalhos Relacionados

A discussão sobre riscos associados à adoção de tecnologias no Poder Judiciário tem ganhado relevo, sobretudo diante da aceleração da transformação digital e da crescente aplicação de inteligência artificial (IA) em tarefas críticas. Os estudos existentes podem ser agrupados em três eixos: riscos no Judiciário brasileiro, riscos técnicos de IA-G e a articulação desses riscos no contexto da governança judicial.

Diversos trabalhos têm destacado vulnerabilidades relacionadas à segurança cibernética e à proteção de dados em tribunais. Alves, Georg e Nunes (2023) mapearam os principais riscos de negócio no ecossistema judiciário, evidenciando que a exposição crescente a ataques cibernéticos e vazamentos de informações compromete a continuidade da prestação jurisdicional, Tabela 2.

Tabela 2 – Riscos de Negócio

N	Descrição
1	Divulgação antecipada de votos, determinações ou decisões
2	Vazamento de informações sigilosas, protegidas por segredo de Justiça ou dados pessoais
3	Emissão ou alteração não autorizada de determinações ou decisões
4	Interrupção da prestação jurisdicional
5	Previsibilidade ou manipulação da distribuição dos processos
6	Perda de informações
7	Parcialidade ou favorecimentos pessoais
8	Assuntos indesejados ou inadequados em determinações e decisões

9	Julgamentos legítimos, porém, com base em elementos adulterados
10	Espionagem

Fonte: Alves, Georg & Nunes (2023)

Almada e Zanatta (2024) ampliam esse debate ao discutir implicações éticas e jurídicas da aplicação de IA em atividades judiciais, apontando riscos de ampliação de desigualdades estruturais. Relatórios do CNJ (2024) mostram que, embora a adoção de IA esteja consolidada em grande parte dos tribunais, os mecanismos de governança e supervisão ainda são incipientes. Mais recentemente, a Resolução CNJ nº 615/2025 estabeleceu diretrizes normativas para o uso responsável de IA no Judiciário, prevendo salvaguardas como supervisão humana, rastreabilidade e auditoria (CNJ, 2025).

No campo mais amplo da ciência da computação, a literatura identifica riscos específicos ou agravados pela adoção de modelos generativos. Chesney e Citron (2019) chamam atenção para os riscos de *deepfakes* e desinformação, com impactos diretos sobre confiança institucional. Wallace et al. (2021) discutem vulnerabilidades ligadas ao envenenamento de dados de treinamento, capazes de comprometer a integridade de modelos em sua origem. Greshake et al. (2023) demonstram como ataques de *prompt injection* podem manipular *outputs* e burlar salvaguardas de sistemas baseados em grandes modelos de linguagem. Arrieta et al. (2020) reforçam a necessidade de explicabilidade e transparência em sistemas de IA e ressaltando que a ausência desses elementos compromete a legitimidade de decisões automatizadas. Complementarmente, Wach et al. (2023) analisam os impactos sociais e econômicos da IA-G, alertando para riscos de toxicidade, viés e homogeneização que afetam diretamente a inclusão e a imparcialidade.

Embora existam avanços significativos tanto na literatura sobre riscos no Judiciário quanto em pesquisas sobre riscos técnicos de IA-G, nota-se a ausência de estudos que integrem essas duas dimensões em um modelo metodológico voltado ao processo judicial. Até o momento, não se identificaram propostas de metodologias que relacionem riscos de negócio do Judiciário com riscos específicos da IA-G em atividades críticas como a elaboração de despachos e decisões. Essa lacuna é particularmente relevante porque tais documentos constituem a base da legitimidade institucional e da confiança pública. O presente estudo se insere nesse espaço, propondo uma metodologia que combina a técnica *bow-tie* e o uso de *heat maps* (mapas de calor) para correlacionar riscos de negócio e riscos de IA-G, oferecendo uma abordagem estruturada para gestores do Judiciário.

Metodologia

A presente pesquisa é de natureza aplicada, pois busca propor e validar uma metodologia prática de avaliação de riscos associados ao uso de IA-G no processo judicial. De acordo com Vergara (2016), pesquisas aplicadas visam não apenas compreender fenômenos, mas propor soluções que possam ser utilizadas em contextos concretos.

Quanto aos objetivos, trata-se de uma investigação de caráter exploratório e descritivo. Segundo Gil (2019), pesquisas exploratórias permitem maior familiaridade com problemas emergentes, sobretudo quando ainda pouco estudados, enquanto as pesquisas

descritivas têm como finalidade apresentar características de determinado fenômeno ou sistematizar relações entre variáveis. Assim, este estudo é exploratório por investigar os riscos decorrentes do uso de IA-G em um contexto ainda em consolidação no Brasil, e descritivo por organizar e correlacionar riscos já identificados na literatura e em documentos institucionais.

No que se refere à abordagem, o estudo adota método qualitativo, uma vez que privilegia a análise interpretativa de documentos, normas e artigos científicos em detrimento de técnicas estatísticas. Creswell (2014) observa que abordagens qualitativas são especialmente adequadas para compreender fenômenos complexos em contextos sociais e institucionais.

Quanto aos procedimentos técnicos, trata-se de uma pesquisa bibliográfica e documental. Foram analisados artigos científicos indexados em bases internacionais, normas técnicas como a ABNT NBR ISO 31000:2018 e o NIST AI Risk Management Framework (2023, 2024), legislações nacionais (como a Resolução CNJ nº 615/2025) e relatórios oficiais do CNJ. No contexto específico de riscos de negócio do Judiciário, foram utilizados dois estudos de referência: (i) *Enhancing cybersecurity in the judiciary: Integrating additional controls into the CIS framework.* (Alves et al., 2025), e (ii) Judiciário sob ataque hacker: Riscos de negócio para segurança e tribunais brasileiros (Alves, Georg & Nunes, 2023). Esses trabalhos forneceram a base empírica para a identificação e classificação dos riscos de negócio correlacionados neste estudo. Não foram aplicados questionários, entrevistas ou observações de campo. Também foram consideradas publicações relevantes da literatura internacional, como Arrieta et al. (2020), Bender et al. (2021) e Brundage et al. (2018), que abordam riscos emergentes, impactos sociais e limites éticos da IA. Além disso, a pesquisa incorporou contribuições atuais da 'Trilha de Implantação de Governança de Dados para IA (Pinto, 2025), que oferecem uma abordagem prática e integradora para a governança de dados e de IA enfatizando os pilares da governança de IA e o papel da ética, da conformidade e da gestão estruturada de riscos.

A análise dos dados foi conduzida em duas etapas complementares. Primeiramente, utilizou-se os riscos de negócio presentes nos estudos de Alves et al. (2023; 2025), Tabela 2, e os riscos de IA-G apontados pelo NIST (2024), Tabela 1. Em seguida, aplicou-se a técnica *bow-tie*, que permite representar graficamente ameaças, barreiras preventivas, consequências e barreiras mitigadoras (CGE *Risk Management Solutions*, 2020). A priorização dos riscos correlacionados foi organizada por meio de *heat maps*, recurso amplamente adotado em gestão de riscos (Hillson, 2009).

Resultados

A aplicação combinada do *heat map*, Figura 1, e da técnica *bow-tie*, Tabela 3, permitiu correlacionar os dez riscos de negócio do processo de despachos e decisões no sistema judiciário brasileiro (Alves, Georg & Nunes, 2023) com os doze riscos de IA-G (NIST.AI.600-1, 2024), evidenciando pontos de exposição e viabilizando a priorização de controles a serem implementados. O *bow-tie*, Tabela 3, estruturou as relações causais entre vulnerabilidades e ameaças da IA-G (causas) e impactos nos riscos de negócio (consequências), exemplificando possíveis barreiras preventivas e mitigadoras às funções e controles recomendados no NIST.AI.600-1 (2024). Em complemento, o *heat map* expôs a densidade de interconexões entre riscos, fornecendo um critério visual para priorização por conectividade e classificação em grupos.

O referido agrupamento de padrões quantitativos reflete diferentes níveis de contribuição na mitigação dos riscos de negócio, com base na conectividade entre os riscos IA-G e os riscos de negócio. Essa divisão permite uma análise mais granular: o Grupo 1, com maior número de cruzamentos; o Grupo 2 e o Grupo 3, com níveis intermediários de conectividade; e o Grupo 4, sem cruzamentos, Tabela 4.

Observa-se que alguns riscos, como Confabulação (RGAI-2), Segurança da Informação (RGAI-6), e Integridade da Informação (RGAI-7), apresentaram elevado número de conexões (seis, seis, e cinco, respectivamente), indicando que atuam como nós centrais de risco capazes de amplificar vulnerabilidades em múltiplas frentes. Já riscos como Privacidade de Dados (RGAI-4), Configuração Humano-IA (RGAI-6), Conteúdo Obsceno ou Abusivo (RGAI-10) e Toxicidade, Viés e Homogeneização (RGAI-11) apresentaram duas conexões cada, posicionando-se como pontos intermediários de exposição que exigem atenção, mas em escopo mais delimitado.

Figura 1 – *Heat Map* (Riscos de Negócio versus Riscos IA-G)

Riscos de Negócio	RGAI-1	RGAI-2	RGAI-3	RGAI-4	RGAI-5	RGAI-6	RGAI-7	RGAI-8	RGAI-9	RGAI-10	RGAI-11	RGAI-12	Total
RNEG-1 Divulgação antecipada de votos, d	0	1	0	0	0	0	1	1	0	0	0	0	3
RNEG-2 Vazamento de informações sigilosa	0	1	0	1	0	0	1	1	0	0	0	0	4
RNEG-3 Emissão ou alteração não autoriza	0	1	0	0	0	0	1	1	0	0	0	0	3
RNEG-4 Interrupção da prestação jurisdicio	0	0	0	0	0	0	0	1	0	0	0	0	1
RNEG-5 Previsibilidade ou manipulação da	0	0	0	0	0	1	0	0	0	0	1	1	3
RNEG-6 Perda de informações	0	1	0	0	0	0	1	1	0	0	0	0	3
RNEG-7 Parcialidade ou favorecimentos pe	0	0	0	1	0	1	0	0	0	1	0	0	3
RNEG-8 Assuntos indesejados ou inadequa	1	1	1	0	0	0	0	0	1	1	1	0	6
RNEG-9 Julgamentos legítimos, porém, con	0	1	0	0	0	0	1	0	0	0	0	0	2
RNEG-10 Espionagem de outras nações e ór	0	0	0	0	0	0	0	1	0	0	0	0	1

Fonte: Elaborado pelos autores

Tabela 3 - Aplicação da Técnica *bow-tie*

N	Causas	Barreiras Preventivas	Evento Central	Consequências	Barreiras Mitigadoras
	Riscos de IA-G	Exemplos de Controles	Riscos de Negócio	Riscos de Negócio	Exemplos de Controles
1	2-Confabulação 7-Integridade da Informação 8-Segurança da Informação	Revisão e verificação de fontes (MS) Validação de informações e fontes (MS, MG) Salvaguardas de integridade, controle de acesso, criptografia (MP, MS) Monitoramento de conteúdo sensível (MG, MS) Segurança cibernética e controle de vulnerabilidades (MS)	1-Divulgação antecipada de votos, determinações ou decisões	2-Vazamento de informações sigilosas 6-Perda de informações 3-Emissão não autorizada	Auditoria e monitoramento (MG, MS) Resposta a incidentes (GV, MG) Revisão contínua de logs e *outputs* (MG)
2	2-Confabulação 4-Privacidade de Dados 7-Integridade da Informação 8-Segurança da Informação	Proteção de dados pessoais (MP, MS) Técnicas de anonimização, controles de acesso (MP, MS) Segmentação de rede (MP) Monitoração contínua Salvaguarda da integridade (MS)	2-Vazamento de informações sigilosas, protegidas por segredo de Justiça ou dados pessoais	7-Parcialidade ou favorecimento 10-Espionagem	Notificação e rastreamento de incidentes (MG) Gestão de crise e contenção de vazamentos (MG, MP) Comunicação à autoridade (GV)
3	2-Confabulação 7-Integridade da Informação 8-Segurança da Informação	Assinatura digital e controle de versões (MS) Validação cruzada IA-humano (MS, MG) Segregação de funções (GV) Monitoração de integridade (MS)	3-Emissão ou alteração não autorizada de determinações ou decisões	9-Decisões com base em elementos adulterados 1-Divulgação antecipada	Auditoria forense (MG) Revisão e correção de registro (MG, MS) Revisão processual (MG)

4	8-Segurança da Informação	Firewalls (MS, MG), backup em tempo real (MS, MG), plano de continuidade (GV, MS), redundância de infraestrutura (MS, MG)	4-Interrupção da prestação jurisdicional	6-Perda de informações 8-Assuntos inadequados	Restauração rápida (MS) Redundância amplificada (MP) Execução de plano de contingência (MG)
5	6-Configuração Humano-IA 1-Toxicidade, Viés e Homogeneização 11-Cadeia de Valor 12-Integração de Componentes	Transparência de algoritmos, interfaces transparentes (GV) Curadoria e revisão independente (MS, MG) Avaliação de fornecedores e transparência da cadeia (GV) Métricas de diversidade (MG)	5-Previsibilidade ou manipulação da distribuição dos processos	7-Parcialidade 3-Emissão não autorizada	Inspeção externa (MG) Reversão de decisões suspeitas (MG) Contestação processual (GV)
6	2-Confabulação 7-Integridade da Informação 8-Segurança da Informação	Backup automático e redundância (MS, MP) Verificação e controle de integridade de dados (MS, MG) Monitoramento de falhas operacionais (MG)	6-Perda de informações	4-Interrupção jurisdicional 2-Vazamento de informações	Recuperação de backup (MG) *Disaster recovery* (MS, MG) Reconstituição parcial de dados (MG)
7	4-Privacidade de Dados 6-Configuração Humano-IA 10-Conteúdo Obsceno, Degradante e/ou Abusivo 11-Toxicidade, Viés e Homogeneização	Curadoria e validação de dados (MS, MG) Monitoramento e detecção de viés (MG) Moderação e filtros automáticos (MS) Revisão humana (GV, MS)	7-Parcialidade ou favorecimentos pessoais	8-Assuntos inadequados 3-Emissão não autorizada	Atualização de *datasets* (MS, MG) Auditoria e contestação processual (GV, MG) Revisão manual de outputs (MG)

8	1-Informações CBRN 3-Recomendações Perigosas ou Violentas 2-Confabulação 9-Propriedade Intelectual 10-Conteúdo Obsceno, Degradante e/ou Abusivo 11-Toxicidade, Viés e Homogeneização	Filtros e detecção de conteúdo crítico (MS, MG) Blacklists/whitelists temáticos (MS, MG) Moderação automática (MS) Detecção de plágio, direitos autorais (MS, GV)	8-Assuntos indesejados ou inadequados em determinações e decisões	1-Divulgação antecipada 7-Parcialidade	Retirada de conteúdo (MG) Revisão e responsabilização disciplinar (GV, MG) Revisão manual (MG)
9	2-Confabulação 7-Integridade da Informação	Controle de versões, registro imutável (MS, MG) Validação cruzada IA-humano (MS) Documentação de prevalência de erros e histórico de alterações (MG)	9-Julgamentos legítimos, porém, com base em elementos adulterados	3-Emissão/alteração não autorizada 6-Perda de informações	Investigação forense de registros (MG) Reprocessamento e auditoria pós-processo (MG, MS)
10	8-Segurança da Informação	Monitoramento de rede e de acesso (MS, MG) Segmentação de sistemas e autenticação forte (MP, MS) Treinamento de pessoal em cibersegurança (GV, MS) Detecção de anomalias (MG)	10-Espionagem de outras nações e/ou grupos de interesse	2-Vazamento de informações 6-Perda de informações	Isolamento de sistemas afetados (MG) Revogação de credenciais/acesso (MS, MG) Resposta a incidentes cibernéticos (MG)

Fonte: Elaborado pelos autores

Tabela 4 – Conexões e Grupos

Riscos de IA-G RGAIs (NIST, 2024)		Conexões RGAI vs RNEGs	Agrupamento
1	Informações CBRN	1 conexão	Grupo 3
2	Confabulação	6 conexões	Grupo 1
3	Recomendações Perigosas ou Violentas	1 conexão	Grupo 3
4	Privacidade de Dados	2 conexões	Grupo 2
5	Impacto Ambiental	0 conexões	Grupo 4
6	Configuração Humano-IA	2 conexões	Grupo 2
7	Integridade da Informação	5 conexões	Grupo 1
8	Segurança da Informação	6 conexões	Grupo 1
9	Propriedade Intelectual	1 conexão	Grupo 3
10	Conteúdo Obsceno, Degradante e/ou Abusivo	2 conexões	Grupo 2
11	Toxicidade, Viés e Homogeneização	2 conexões	Grupo 2
12	Cadeia de Valor e Integração de Componentes	1 conexão	Grupo 3

Fonte: Elaborado pelos autores

Os riscos Informações CBRN (RGAI-1), Recomendações Perigosas ou Violentas (RGAI-3), Propriedade Intelectual (RGAI-9) e Cadeia de Valor e Integração de Componentes (RGAI-12) apresentam uma conexão. Em contraste, riscos como Impacto Ambiental (RGAI-5) não apresentaram conexões diretas com os riscos de negócio, formando um grupo isolado (Grupo 4) com menor prioridade imediata. Essa distribuição evidencia que os Grupos 1 e 2, que concentram os riscos mais conectados, devem receber prioridade na aplicação de controles e salvaguardas, uma vez que sua mitigação tende a gerar efeitos de proteção sistêmica sobre a cadeia de riscos do processo judicial.

Discussões

Os resultados obtidos evidenciam que a aplicação da técnica *bow-tie* associada ao *heat map* foi eficaz para correlacionar riscos de negócio do Poder Judiciário com riscos específicos de IA-G. Essa abordagem revelou que determinados riscos atuam como nós centrais de

vulnerabilidade, capazes de amplificar impactos em diferentes dimensões do processo judicial. Casos como a confabulação (RGAI-2), a integridade da informação (RGAI-7) e a segurança da informação (RGAI-8) confirmaram a percepção de que a introdução de IA-G não substitui, mas intensifica, preocupações já mapeadas no campo da segurança cibernética e da confiabilidade institucional (Alves, Georg & Nunes, 2023; NIST, 2024).

Um primeiro ponto de reflexão refere-se ao alinhamento entre riscos técnicos e riscos de negócio. A análise demonstrou que as ameaças geradas pela IA-G não se restringem ao âmbito tecnológico, mas possuem repercussões diretas na legitimidade institucional, na imparcialidade e na confiança pública. Tal constatação converge com Almada e Zanatta (2024), que alertam para os riscos éticos e jurídicos da automação de decisões judiciais, e reforça que os riscos de IA-G devem ser tratados como questões de governança e não apenas de infraestrutura de TI.

Em segundo lugar, os achados apontam para a necessidade de priorização por conectividade. Os riscos mais interligados (Grupos 1 e 2) demandam atenção imediata, já que sua mitigação pode gerar efeitos sistêmicos na cadeia de riscos do processo judicial. Isso indica que estratégias de gestão baseadas em pacotes integrados de controles são mais eficazes do que medidas isoladas, corroborando a literatura sobre gestão de riscos emergentes (ABNT, 2018; ABNT ISO/TS 31050, 2025).

Outro aspecto relevante é a relação entre explicabilidade, supervisão humana e legitimidade institucional. O risco identificado como “assuntos indesejados ou inadequados em determinações e decisões” (RNEG-8) ilustra a vulnerabilidade a conteúdos tóxicos, enviesados ou fabricados. Essa constatação reforça a importância de mecanismos de explicabilidade (Arrieta et al., 2020) e de supervisão humana qualificada (CNJ, 2025) como barreiras indispensáveis, sob pena de se comprometer a confiança social no processo judicial.

Por fim, a análise evidencia que há áreas de risco negligenciadas. O “Impacto Ambiental” (RGAI-5), por exemplo, não apresentou conexões com os riscos de negócio nesta amostra, mas a literatura aponta para a crescente relevância da sustentabilidade digital (Gartner, 2024; ENISA, 2023). Embora hoje não se configure como prioridade imediata, trata-se de um vetor que deve ser monitorado prospectivamente, em especial considerando o aumento da pegada de carbono associada a modelos de larga escala.

Portanto, os resultados deste estudo não apenas validam a utilidade da metodologia proposta, mas também destacam que a adoção responsável de IA-G no Judiciário requer uma abordagem integrada que combine segurança cibernética, governança de dados, ética e supervisão humana. Essa visão sistêmica fortalece a resiliência institucional e pode servir de referência para outros setores da administração pública que enfrentam desafios semelhantes.

Conclusões e Recomendações

O presente estudo buscou analisar os riscos associados ao uso da IA-G no processo de elaboração de despachos e decisões judiciais, propondo uma metodologia de avaliação que integra a técnica *bow-tie* com o uso de *heat maps*. A escolha por correlacionar riscos de negócio já identificados em estudos anteriores com os riscos de IA-G sistematizados pelo NIST (2024) permitiu demonstrar não apenas as ameaças técnicas, mas também suas repercussões no plano institucional, ético e organizacional do Poder Judiciário. Ao adotar uma perspectiva que articula riscos tecnológicos e de negócio, o estudo contribui para preencher uma lacuna na literatura nacional e internacional, que trata em sua maioria, essas dimensões de forma isolada.

Os resultados alcançados indicaram que determinados riscos ocupam posição central e apresentam maior potencial de amplificação de vulnerabilidades, como a confabulação, a integridade da informação e a segurança da informação. Outros, por sua vez, mostraram-se igualmente relevantes por sua multiplicidade de conexões, como os relacionados à governança de conteúdo e à configuração humano-IA, revelando que a complexidade do fenômeno exige não apenas barreiras tecnológicas, mas também mecanismos institucionais de supervisão, explicabilidade e controle humano qualificado. A priorização por conectividade, viabilizada pelo uso do *heat map*, demonstrou ser um recurso poderoso para orientar a alocação de esforços de mitigação, sugerindo que a gestão de riscos em ambientes judiciais deve adotar pacotes integrados de controles, capazes de produzir efeitos sistêmicos mais robustos do que medidas isoladas e fragmentadas.

Esses achados reforçam a importância de compreender a IA-G não apenas como um recurso tecnológico a ser incorporado pela infraestrutura de TI do Judiciário, mas como um elemento que afeta diretamente a legitimidade institucional e a confiança pública nas decisões judiciais. A integração entre segurança cibernética, governança de dados, ética e mecanismos de supervisão humana aparece como condição para garantir que os benefícios da IA-G, como maior eficiência, agilidade e acessibilidade, não sejam impactados por riscos que comprometam direitos fundamentais ou a confiançado sistema de justiça. Nesse sentido, a metodologia aqui proposta pode orientar gestores do Judiciário e inspirar outros setores da administração pública que enfrentam desafios semelhantes no uso de tecnologias emergentes.

De todo modo, é importante reconhecer as limitações deste estudo. Por tratar-se de uma pesquisa qualitativa e documental, baseada em uma amostra não probabilística e por conveniência, composta por normas, relatórios e artigos científicos, a generalização dos resultados permanece restrita. Além disso, a análise se concentrou na etapa de elaboração de despachos e decisões judiciais, deixando de fora outras fases processuais igualmente

impactadas pelo uso de IA-G, como a triagem, a classificação de casos e a produção de minutas. Ademais, a própria técnica *bow-tie*, ainda que útil para mapear relações causais, depende da interpretação dos pesquisadores na definição das conexões e barreiras, o que introduz vieses na análise.

Apesar dessas restrições, entende-se que a metodologia poderá ser validada empiricamente em projetos-piloto em tribunais brasileiros, ampliando sua robustez por meio da combinação com métodos quantitativos, como simulações de cenários, análise multicritério ou modelagem estatística. Outra frente de avanço consiste em expandir o escopo da análise para outras etapas do processo judicial e até mesmo para diferentes áreas da administração pública, de modo a testar a aplicabilidade da abordagem em contextos variados. Além disso, pesquisas que integrem a percepção de magistrados, servidores e especialistas em tecnologia poderão enriquecer a compreensão dos riscos e controles, trazendo ao debate a perspectiva prática daqueles que operam diretamente no ecossistema judicial.

Em síntese, este estudo reforça que a adoção da IA-G no Poder Judiciário não pode ser tratada como mera inovação tecnológica, mas como uma transformação institucional que envolve riscos significativos, cuja gestão deve ser pautada por metodologias sólidas, integradas e replicáveis. A contribuição aqui apresentada reside justamente na proposição de uma abordagem que alia rigor técnico e aplicabilidade prática, oferecendo aos gestores públicos instrumentos para alinhar inovação e governança responsável, de modo a garantir que a modernização tecnológica do Judiciário se traduza em ganhos efetivos de eficiência, legitimidade e confiança social.

Referências

Alves, R. S., Georg, M. A., & Nunes, R. R. (2023). Judiciário sob ataque hacker: Riscos de negócio para segurança cibernética em tribunais brasileiros. Revista Ibérica de Sistemas e Tecnologias de Informação, 2023(1), 344-357

Alves, R. S., da Silva, J. P. B., Ribeiro Junior, L. A., & Nunes, R. R. (2025). Enhancing cybersecurity in the judiciary: Integrating additional controls into the CIS framework. Computers & Security, 157, 104584. https://doi.org/10.1016/j.cose.2025.104584

Almada, M., & Zanatta, R. A. F. (2024). Inteligência artificial, direito e pesquisa jurídica. Revista USP, (141), 51–64. https://doi.org/10.11606/issn.2316-9036.i141p51-64

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies,

opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Associação Brasileira de Normas Técnicas. (2018). *ABNT NBR ISO 31000:2018 – Gestão de riscos: Diretrizes*. ABNT.

Associação Brasileira de Normas Técnicas. (2023). *ABNT NBR ISO/IEC 23894:2023 – Tecnologia da informação: Inteligência artificial – Orientações sobre gestão de riscos*. ABNT.

Associação Brasileira de Normas Técnicas. (2023). *ABNT NBR ISO/IEC 22989:2023 – Tecnologia da informação: Inteligência artificial – Conceitos e terminologia*. ABNT.

Associação Brasileira de Normas Técnicas. (2024). *ABNT NBR ISO/IEC 42001:2024 – Tecnologia da informação: Inteligência artificial – Sistema de gestão*. ABNT.

Associação Brasileira de Normas Técnicas. (2025). *ABNT ISO/TS 31050:2025 – Gestão de riscos: Diretrizes para riscos emergentes e resiliência*. ABNT.

Brasil. Decreto nº 9.203, de 22 de novembro de 2017. Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional. Diário Oficial da União.

Brasil. (2021). Estratégia Brasileira de Inteligência Artificial (EBIA). Ministério da Ciência, Tecnologia e Inovações. https://www.gov.br/mcti/pt-br/assuntos/noticias/ebia-lanca-estrategia-brasileira-de-inteligencia-artificial

CGE Risk Management Solutions. (2020). *BowTie methodology: A practical guide*. CGE Risk.

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

Conselho Nacional de Justiça. (2021). Justiça 4.0. Acesso em 21 de maio de 2022. https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/

Conselho Nacional de Justiça. (2024). Programa Justiça 4.0 divulga resultados de pesquisa sobre IA no Judiciário brasileiro. Acesso em 08 de setembro de 2024. https://www.cnj.jus.br/programa-justica-4-0-divulga-resultados-de-pesquisa-sobre-ia-no-judiciario-brasileiro/

Conselho Nacional de Justiça. (2025). *Resolução nº 615, de 11 de março de 2025. Dispõe sobre a governança e o uso de inteligência artificial no âmbito do Poder Judiciário*. Brasília: CNJ.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

European Union Agency for Cybersecurity (ENISA). (2023). AI Cybersecurity: Towards Trustworthy AI Development. https://www.enisa.europa.eu/publications/ai-cybersecurity-towards-trustworthy-ai-development

Gartner. (2024). O uso indevido da IA generativa diminui o valor da IA nas organizações. Acesso em 08 de setembro de 2024. https://www.gartner.com.br/pt-br/artigos/quando-nao-usar-a-ia-generativa

Gil, A. C. (2019). *Métodos e técnicas de pesquisa social* (7ª ed.). Atlas.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). More than you've asked for: A comprehensive analysis of indirect prompt injection attacks in LLM applications. *Proceedings of the 32nd USENIX Security Symposium*. https://arxiv.org/abs/2302.12173

Hino, M. C., & Cunha, M. A. (2020). Adoção de tecnologias na perspectiva de profissionais de direito. Revista Direito GV, 16(1), e1952. https://doi.org/10.1590/2317-6172201952

Hillson, D. (2009). *Managing risk in projects*. Gower.

National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI-600-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI-600-1

Russel, S., & Norvig, P. (2010). Artificial Intelligence: a modern approach. Pearson Education

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (pp. 3-17). https://arxiv.org/abs/1610.05820

Supremo Tribunal Federal. (2023). *STF conclui chamamento público para uso de inteligência artificial*. https://noticias.stf.jus.br/postsnoticias/stf-conclui-chamamento-publico-para-uso-de-inteligencia-artificial/

UNESCO. (2024). AI and the Rule of Law: Capacity Building for Judicial Systems. Acesso em 08 de setembro de 2024. https://www.unesco.org/en/artificial-intelligence/rule-law/mooc-judges

Vergara, S. C. (2016). *Projetos e relatórios de pesquisa em administração* (16ª ed.). Atlas.

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., ... & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Business and Economics Review, 11(2), 7-30

Wallace, E., Feng, S., Kandpal, N., Gardner, M., & Singh, S. (2021). Universal adversarial triggers for attacking and analyzing NLP. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2153–2169. https://doi.org/10.18653/v1/2021.naacl-main.168