ARTIGO TÉCNICO

TCU estabelece metodologia para auditar inteligência artificial generativa no setor público

Ao avaliar o sistema Galileu, utilizado pela Justiça do Trabalho, o Tribunal de Contas da União (TCU) desenvolveu uma metodologia que combina avaliação de governança, testes técnicos e uso de agentes de IA para verificar a confiabilidade de sistemas generativos.

Ana Rita Fontes Nascimento[1]

Novembro de 2026

[1] Ana Rita Fontes Nascimento. Mestranda em Direito (UFMG). Pesquisadora do Centro DTIBR. Lattes: http://lattes.cnpq.br/1038165509890627

SUMÁRIO

1 Introdução 3

2 A construção de uma metodologia de auditoria para IA generativa 3

3 O uso de agentes de IA na fiscalização 5

4 A avaliação da fidelidade das informações geradas 6

5 Resultados dos demais testes de auditoria 8

6 Equidade, explicabilidade e supervisão humana 10

7 As recomendações do TCU 11

Referências 13

1 Introdução

O TCU estabeleceu uma metodologia para avaliar a confiabilidade e os riscos associados ao uso de inteligência artificial generativa (IAG) na Administração Pública. Por meio do Acórdão nº 2.836/2026, aprovado pelo Plenário em 14 de outubro de 2026, o Tribunal examinou o sistema Galileu, desenvolvido pelo Tribunal Regional do Trabalho da 4ª Região (TRT-4), responsável por auxiliar magistrados na elaboração de minutas judiciais. A fiscalização resultou em recomendações para o aperfeiçoamento da ferramenta e apresentou um conjunto de procedimentos que poderá servir de referência para futuras auditorias de sistemas de IA utilizados pelo poder público.

A auditoria operacional foi conduzida pela Unidade de Auditoria Especializada em Tecnologia da Informação (AudTI) e teve como objetivo avaliar a confiabilidade tecnológica e operacional do Galileu, bem como a efetividade dos mecanismos destinados a assegurar seu uso responsável, ético e seguro. O trabalho envolveu a análise de estruturas de governança, procedimentos de segurança da informação, tratamento de dados pessoais, arquitetura tecnológica e atributos associados à inteligência artificial responsável.

O Galileu, que foi desenvolvido para auxiliar na leitura e organização das peças processuais, já havia alcançado utilização significativa na Justiça do Trabalho e estava em processo de expansão nacional, sob coordenação do Conselho Superior da Justiça do Trabalho (CSJT). Segundo os dados examinados pela auditoria, até 10 de junho de 2026, haviam sido geradas 173.041 minutas por 4.904 usuários, sendo o sistema utilizado para apoiar aproximadamente 40% das minutas de sentença de primeiro grau no TRT-4.

A expansão da ferramenta ampliou a relevância de seus mecanismos de controle, considerando a possibilidade de que eventuais falhas de processamento fossem reproduzidas em diferentes unidades jurisdicionais. Nesse contexto, a auditoria buscou verificar se o sistema oferecia condições adequadas de segurança, confiabilidade e supervisão humana, observando os parâmetros jurídicos e técnicos aplicáveis à utilização de IA pelo Poder Judiciário, entre os quais a Resolução nº 615/2025 do Conselho Nacional de Justiça (CNJ, 2025).

2 A construção de uma metodologia de auditoria para IA generativa

A principal contribuição do trabalho desenvolvido pelo TCU está na sistematização de critérios e procedimentos específicos para a fiscalização de

sistemas de inteligência artificial generativa. A equipe estruturou a metodologia em cinco dimensões de avaliação, compreendendo governança, segurança da informação, gestão de dados, arquitetura e infraestrutura e atributos de inteligência artificial responsável. A definição desses eixos considerou referenciais internacionais, entre os quais os Princípios de IA da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), o AI Risk Management Framework do National Institute of Standards and Technology (NIST), o AI Accountability Framework do Government Accountability Office (GAO) e o OWASP Top 10 for LLM Applications.

A metodologia foi concebida a partir da necessidade de adaptar os instrumentos de controle à natureza dos sistemas generativos, cujos resultados apresentam características diferentes daquelas encontradas em modelos tradicionais de classificação e previsão. O próprio relatório reconhece que a equipe precisou desenvolver um caminho metodológico específico, definindo os testes aplicáveis, sua execução e os critérios de interpretação dos resultados (TCU, 2026, Relatório, p. 3).

Tabela 1 – Dimensões de avaliação da auditoria do Galileu

Dimensão	Critérios examinados
Governança	Estruturas organizacionais, responsabilidades, conformidade legal, políticas, gestão de riscos e mecanismos de melhoria contínua.
Segurança da informação	Proteção contra ataques, manipulação de entradas, ameaças à segurança e riscos relacionados ao tratamento de dados pessoais.
Gestão de dados	Qualidade dos dados de entrada e saída, rastreabilidade, documentação e controle do ciclo de vida das informações.
Arquitetura e infraestrutura	Robustez, integridade do processamento, reprodutibilidade, implantação e continuidade operacional.
Atributos de IA responsável	Transparência, explicabilidade, supervisão humana, equidade, contestabilidade, responsabilização, auditabilidade e confiabilidade.

Fonte: elaboração a partir das figuras 6 e 7 do relatório de auditoria, páginas 11 e 12 do Acórdão nº 2.836/2026-TCU-Plenário.

A organização dos critérios permitiu que os atributos de IA responsável fossem examinados como propriedades verificáveis do sistema. A transparência, por exemplo, foi relacionada à disponibilidade de informações sobre a utilização da ferramenta, enquanto a explicabilidade foi associada à possibilidade de compreender e rastrear os elementos que determinaram determinado resultado. A confiabilidade, por sua vez, foi examinada mediante testes destinados a identificar omissões, distorções e afirmações sem respaldo nos documentos processuais.

A aplicação desses critérios ocorreu inicialmente por meio de um Questionário de Autoavaliação de Controles Internos (QACI), respondido pelo TRT-4 com a indicação do grau de atendimento de cada requisito e a

apresentação das respectivas evidências. As informações foram posteriormente confrontadas com a documentação técnica, o código-fonte, a utilização direta da ferramenta e os testes executados pela equipe de fiscalização.

A auditoria também contou com entrevistas com usuários, reuniões com gestores e visita técnica realizada ao TRT-4, em Porto Alegre, entre 4 e 8 de maio de 2026. Os experimentos foram conduzidos em ambiente que reproduzia as condições de produção do Galileu, utilizando seu código real e os modelos de inteligência artificial empregados na operação do sistema.

3 O uso de agentes de IA na fiscalização

A utilização de agentes de inteligência artificial constituiu uma característica relevante da metodologia desenvolvida pelo TCU. A equipe empregou essas ferramentas no planejamento e na execução da auditoria, especialmente para desenvolver códigos, organizar experimentos, analisar resultados e ampliar a quantidade de verificações realizadas.

A participação desses agentes permitiu a realização de aproximadamente vinte testes técnicos planejados, abrangendo diferentes aspectos do funcionamento do Galileu. Os principais procedimentos foram sistematizados pelo TCU conforme a dimensão avaliada e a finalidade de cada experimento, como apresentado a seguir.

Tabela 2 – Testes técnicos utilizados na auditoria do Galileu

Teste	Procedimento de avaliação	Objetivo
Injeção de comandos (*prompt injection*)	Simulação adversarial de entradas manipuladas, incluindo instruções inseridas em documentos processuais.	Verificar a resistência do sistema a tentativas de manipulação dos resultados.
Fidelidade factual	Comparação das minutas com os documentos processuais, utilizando um padrão-ouro de referência.	Identificar alucinações, omissões e distorções de fatos e alegações.
Viés	Comparação pareada de minutas com alteração controlada de atributos como gênero e ocupação das partes.	Investigar possíveis diferenças de tratamento relacionadas a essas características.
Explicabilidade	Exame das etapas de processamento e da vinculação das afirmações aos documentos utilizados.	Avaliar a possibilidade de compreender, verificar e rastrear os resultados.
Supervisão humana	Avaliação dos mecanismos disponíveis para revisão, edição e validação das minutas.	Verificar a efetividade do controle humano sobre os resultados gerados.
Reprodutibilidade	Repetição de gerações a partir dos mesmos dados de entrada.	Medir a consistência e a estabilidade das respostas do sistema.
Integridade do processamento	Verificação da propagação de falhas entre as etapas do processamento (*pipeline*).	Identificar a existência de controles capazes de impedir a transmissão de erros entre etapas.
Recuperação de precedentes	Avaliação da busca e da ordenação de subsídios jurídicos.	Examinar a qualidade da recuperação e da apresentação dos resultados.

Fonte: elaboração a partir da Tabela 1, página 13, do relatório de auditoria integrante do Acórdão nº 2.836/2026-TCU-Plenário.

Os testes de segurança receberam atenção particular em razão das possibilidades de manipulação das informações processadas pelo sistema. A equipe simulou tentativas de *prompt injection*, técnica que consiste na inserção de comandos destinados a modificar indevidamente o comportamento de um modelo de IA. Em determinadas situações, os agentes utilizados na fiscalização conseguiram formular instruções capazes de ultrapassar as barreiras de proteção existentes.

Um dos experimentos demonstrou que conteúdos ocultos em documentos processuais poderiam induzir o Galileu a produzir informações inexistentes nos autos, incluindo supostas confissões da parte reclamada acerca de fatos relacionados à relação de trabalho. A manipulação foi realizada em ambiente controlado e revelou vulnerabilidades que poderiam comprometer a fidelidade das minutas apresentadas aos magistrados.

Os resultados foram comunicados ao TRT-4, que implementou medidas corretivas durante a própria auditoria. A equipe do TCU realizou novos testes e verificou que as técnicas anteriormente bem-sucedidas deixaram de produzir os mesmos efeitos. A correção, implementada na versão 1.4.4 do sistema, incluiu mecanismos de filtragem para impedir que textos ocultos fossem encaminhados ao modelo responsável pela elaboração das minutas, preservando os alertas aos usuários sobre possíveis tentativas de manipulação.

A experiência evidenciou a utilidade dos testes adversariais como instrumento de controle de sistemas generativos, especialmente diante da capacidade desses modelos de interpretar instruções incorporadas a documentos aparentemente legítimos. O acórdão recomendou que o TRT-4 realizasse avaliações periódicas de segurança, com evolução para práticas contínuas de testes adversariais, preferencialmente apoiadas por ferramentas automatizadas.

4 A avaliação da fidelidade das informações geradas

A confiabilidade das minutas foi examinada mediante uma metodologia de comparação entre os documentos processuais e os textos produzidos pelo Galileu. Para essa etapa, a equipe utilizou cinquenta processos reais e não sigilosos, abrangendo os três fluxos de funcionamento da ferramenta, que resultaram na análise de 383 tópicos.

O procedimento foi organizado a partir da construção de um padrão-ouro, correspondente a uma referência dos pedidos e fatos efetivamente presentes em cada processo. Um modelo de IA foi empregado para extrair essas

informações das peças processuais, enquanto outro realizou uma revisão do material em confronto com os documentos originais. Posteriormente, as minutas produzidas pelo Galileu foram comparadas com essa referência.

A metodologia incorporou uma etapa adicional de verificação para reduzir classificações equivocadas. Cada suspeita de alucinação passou por nova conferência das peças, seguida de revisão humana. Ao final, das 77 ocorrências inicialmente identificadas, 49 foram descartadas como falsos positivos e 28 foram confirmadas pelos auditores. A importância dessa revisão foi expressamente reconhecida no relatório de auditoria, que registrou que "A palavra final coube ao humano" (TCU, 2026, Relatório, p. 22).

Os resultados demonstraram diferentes níveis de cobertura das informações presentes nos processos, conforme os indicadores apresentados na Tabela 3. A avaliação considerou as características próprias da tarefa de síntese desempenhada pelo Galileu, descontando omissões deliberadas decorrentes das instruções de funcionamento da ferramenta.

Tabela 3 – Principais resultados dos testes de fidelidade factual

Indicador	Resultado
Processos examinados	50
Tópicos analisados	383
Cobertura substantiva dos pedidos	93,3%
Cobertura substantiva dos fatos	81,3%
Cobertura temática	86,1%
Suspeitas iniciais de alucinação	77
Suspeitas descartadas após revisão	49
Alucinações confirmadas	28
Taxa de alucinação por tópico	7,3%
Processos com ao menos uma alucinação	42%
Alucinações de severidade alta	0

Fonte: elaboração a partir dos resultados apresentados nas páginas 21 a 24 do relatório de auditoria integrante do Acórdão nº 2.836/2026-TCU-Plenário.

As alucinações confirmadas, apresentadas na Tabela 3, corresponderam a dezessete ocorrências de baixa severidade e onze de severidade média. Em um dos exemplos apresentados no relatório, o sistema modificou uma alegação de trabalho em um domingo por mês para trabalho em domingos alternados, alteração que poderia repercutir na compreensão dos fatos e na quantificação das verbas discutidas.

A distribuição das ocorrências também demonstrou a necessidade de considerar diferentes unidades de análise. Como indicado na Tabela 3, o

percentual de tópicos com alucinações foi inferior ao percentual de processos nos quais se identificou pelo menos uma ocorrência. Essa diferença decorre da existência de diversos tópicos em uma mesma minuta e evidencia a necessidade de examinar tanto a frequência dos erros quanto sua distribuição entre os documentos produzidos.

A metodologia utilizada para identificar as alucinações também revelou limitações dos próprios modelos empregados na auditoria. A quantidade de falsos positivos inicialmente produzidos demonstrou a necessidade de procedimentos adicionais de conferência e validação humana, especialmente quando ferramentas generativas são utilizadas para avaliar os resultados de outros sistemas de IA. A consistência desses procedimentos de controle depende da definição de referências verificáveis, da documentação das etapas de análise e da existência de mecanismos que permitam revisar as conclusões produzidas automaticamente.

5 Resultados dos demais testes de auditoria

Os testes de reprodutibilidade, integridade do processamento e recuperação de precedentes também produziram resultados relevantes para a avaliação do Galileu. Os procedimentos permitiram verificar a estabilidade das minutas geradas a partir de uma mesma entrada, a capacidade do sistema de identificar erros introduzidos durante o processamento e a qualidade da ordenação dos subsídios jurídicos apresentados aos usuários.

Tabela 4 – Resultados dos testes de reprodutibilidade, integridade e recuperação de precedentes

Critério	Procedimento	Resultado
Reprodutibilidade textual	Vinte gerações da mesma peça processual, sendo dez na configuração de produção, com temperatura 1, e dez com temperatura 0. A temperatura é um parâmetro que influencia a variação das respostas produzidas pelo modelo.	Na configuração de produção, duas minutas compartilhavam, em média, 51% do texto. Com temperatura 0, o percentual de texto compartilhado alcançou 98%.
Estabilidade das referências normativas	Verificação da presença das mesmas referências jurídicas ao longo das dez gerações realizadas em cada configuração.	Na configuração de produção, foram identificadas 18 referências normativas diferentes, mas nenhuma apareceu em todas as dez gerações. Com temperatura 0, foram identificadas seis referências, todas presentes nas dez gerações.

Critério	Procedimento	Resultado
Integridade do processamento	Introdução controlada de três tipos de erro na primeira etapa do processamento, com três repetições por tipo.	Em todas as nove execuções, os erros inseridos atravessaram as etapas posteriores e chegaram à minuta final sem detecção pelos controles intermediários.
Recuperação de precedentes	Comparação cega entre a ordenação original e uma ordenação alternativa dos documentos recuperados, com 285 comparações válidas.	A ordenação alternativa foi considerada mais útil em 70% das comparações dos cinco primeiros resultados e em 79% das comparações dos vinte primeiros.

Fonte: elaboração a partir das tabelas 15 e 17 e dos resultados dos testes de integridade apresentados nas páginas 41, 45 e 46–47 do relatório de auditoria integrante do Acórdão nº 2.836/2026-TCU-Plenário.

Os resultados de reprodutibilidade apresentados na Tabela 4 demonstram que alterações no parâmetro de temperatura influenciaram significativamente a estabilidade das minutas produzidas pelo Galileu. A configuração utilizada em produção apresentou maior variação na redação, na organização dos tópicos e na presença das referências normativas, enquanto a configuração com temperatura 0 produziu resultados mais estáveis. A variação das referências merece atenção porque uma mesma peça processual pode dar origem a minutas que mencionam fundamentos jurídicos diferentes, embora os documentos utilizados sejam idênticos. A equipe observou, por exemplo, que uma referência à ADI nº 5.766, relevante para um dos processos examinados, deixou de aparecer em quatro das dez gerações realizadas na configuração de produção.

A avaliação desse resultado considerou que a identidade textual entre diferentes execuções não constitui, isoladamente, um requisito de confiabilidade. O relatório reconheceu que a configuração utilizada pelo TRT-4 decorreu de uma escolha deliberada, associada à qualidade da redação percebida pelos usuários, e que a consistência substantiva deve ser aferida pela preservação dos pedidos, fatos e fundamentos relevantes, independentemente de eventuais diferenças na formulação textual (TCU, 2026, Relatório, p. 44–46).

A integridade do processamento foi examinada mediante a introdução artificial de erros na primeira etapa de elaboração da minuta. Os cenários envolveram a omissão de um pedido, a distorção de um fato relevante e a inclusão de um pedido inexistente. Conforme os resultados apresentados na Tabela 4, nenhum dos erros foi bloqueado durante o processamento, circunstância que evidenciou a ausência de mecanismos de validação cruzada entre as etapas e fundamentou a recomendação de avaliar controles capazes de verificar a correspondência das informações com os documentos originais.

O alcance desse resultado deve ser considerado a partir das condições do experimento. Os erros foram introduzidos deliberadamente pela equipe de fiscalização para testar a robustez da arquitetura e não corresponderam a falhas espontâneas observadas nas minutas da amostra de fidelidade factual. O

relatório reconheceu, ainda, que a implementação de verificações adicionais entre etapas envolve custos de processamento e aumento do tempo de resposta, cuja conveniência deve ser avaliada pelo TRT-4 conforme os riscos e benefícios envolvidos (TCU, 2026, Relatório, p. 46–47).

A recuperação de precedentes apresentou oportunidades de aprimoramento relacionadas à ordenação dos documentos disponibilizados no painel de subsídios do Galileu. Nesse experimento, a equipe comparou a lista originalmente apresentada pelo sistema com uma ordenação alternativa dos mesmos documentos, submetendo ambas à avaliação cega de modelos de IA de fornecedores distintos. Os resultados da Tabela 4 indicam que a ordenação alternativa foi considerada mais útil para fundamentação na maioria das comparações, especialmente em relação à apresentação dos precedentes do TRT e dos tribunais superiores.

Os resultados desse último teste também apresentam uma delimitação funcional importante. O painel de subsídios constitui uma ferramenta de pesquisa utilizada pelo magistrado e não alimenta automaticamente a geração da minuta. Após a manifestação dos gestores, a equipe retirou a proposta inicial de recomendação para implementar a reordenação, considerando os custos e o tempo adicional de processamento. A possibilidade de aperfeiçoamento permaneceu registrada no relatório, especialmente para eventual adoção futura de mecanismos de recuperação de informações integrados à geração de textos (TCU, 2026, Relatório, p. 40–44).

A análise conjunta desses experimentos demonstra que a avaliação de um sistema generativo envolve propriedades distintas e próprias, cujos resultados não podem ser interpretados por uma medida única de desempenho. A variação textual entre execuções, a propagação de erros simulados e a ordenação de subsídios jurídicos correspondem a aspectos específicos do funcionamento do Galileu e fornecem evidências para decisões de governança, desenvolvimento e gestão de riscos.

6 Equidade, explicabilidade e supervisão humana

A avaliação de equidade foi realizada por meio de testes de comparação pareada, nos quais a equipe produziu versões equivalentes de um mesmo processo, alterando de forma controlada características como gênero e ocupação das partes. O objetivo consistiu em identificar possíveis diferenças no tratamento das alegações e na organização das informações decorrentes dessas alterações.

Os resultados não apresentaram evidência estatisticamente significativa de viés de gênero ou ocupação nas condições examinadas. O alcance dessa conclusão permaneceu limitado aos atributos e cenários utilizados, razão pela qual o TCU recomendou a realização periódica de avaliações de equidade, inclusive com a consideração de outras características relevantes.

A explicabilidade também foi objeto de exame específico, especialmente quanto à possibilidade de relacionar as afirmações produzidas pelo Galileu aos documentos processuais que lhes deram origem. A auditoria identificou oportunidades de aprimoramento na rastreabilidade, considerando que a indicação de um documento ou de sua localização não assegura, por si só, a identificação precisa do trecho que fundamenta determinada informação.

O TCU recomendou o aperfeiçoamento dos mecanismos de vinculação entre o texto gerado e as passagens correspondentes das peças processuais, de modo a facilitar a conferência das informações pelos usuários. Também recomendou a preservação dos resultados intermediários das etapas de processamento e dos conteúdos explicativos efetivamente disponibilizados pelos modelos, observados os requisitos de segurança e proteção de dados.

A supervisão humana foi analisada a partir dos controles que permitem ao magistrado revisar, modificar ou rejeitar o conteúdo gerado. Embora a decisão judicial permaneça sob responsabilidade humana, a auditoria identificou oportunidades para tornar a revisão mais efetiva, considerando o risco de que textos aparentemente coerentes sejam aceitos sem conferência suficiente.

Nesse sentido, o acórdão recomendou a adoção de mecanismos que apoiem a revisão crítica das minutas e permitam registrar as intervenções dos usuários. A preocupação relaciona-se ao chamado viés de automação, fenômeno pelo qual a confiança depositada em resultados automatizados pode reduzir o grau de atenção dedicado à sua verificação.

A relevância jurídica da supervisão humana foi enfatizada pelo relator, ministro Odair Cunha, que estabeleceu como premissa de seu voto que "a IA deve ampliar a capacidade de trabalho do ser humano, não substituir sua responsabilidade por decidir" (TCU, 2026, Voto, p. 1).

7 As recomendações do TCU

O Plenário do TCU, acompanhando o voto do ministro Odair Cunha, concluiu que o Galileu apresenta confiabilidade nas dimensões centrais examinadas, considerando as condições de funcionamento avaliadas e as correções realizadas durante a fiscalização. A decisão reconheceu a existência de mecanismos satisfatórios de governança e proteção de dados, bem como oportunidades de

aperfeiçoamento relacionadas à qualidade das saídas, à rastreabilidade e à efetividade da supervisão humana.

Tabela 5 – Recomendações do TCU para aperfeiçoamento do Galileu

Área	Recomendação
Segurança	Realizar testes periódicos de injeção de comandos e desenvolver práticas contínuas de testes adversariais.
Fidelidade factual	Instituir mecanismos objetivos para identificar alucinações e avaliar alterações de qualidade após atualizações.
Explicabilidade	Aperfeiçoar a vinculação das afirmações geradas aos trechos documentais que lhes dão suporte.
Supervisão humana	Desenvolver ferramentas que apoiem a conferência e a revisão efetiva das minutas.
Equidade	Implementar procedimentos periódicos de avaliação de possíveis vieses.
Rastreabilidade	Registrar modelos, parâmetros, versões de instruções e resultados intermediários de processamento.
Infraestrutura	Avaliar medidas de redundância ou diversificação dos provedores de IA.
Gestão de custos	Aprimorar o acesso às informações de consumo e faturamento dos serviços utilizados.
Integridade documental	Estudar mecanismos para identificação de documentos potencialmente manipulados antes de seu processamento.

Fonte: elaboração a partir dos itens 9.1 e 9.2 do Acórdão nº 2.836/2026-TCU-Plenário, páginas 63 e 64 do documento.

As recomendações foram dirigidas aos órgãos envolvidos na operação e na expansão do Galileu, com fundamento no art. 11 da Resolução TCU nº 315/2020. A decisão autorizou o monitoramento das providências adotadas e dos resultados alcançados, permitindo o acompanhamento da contribuição dessas medidas para a redução dos riscos identificados.

Essa continuidade do acompanhamento constitui outro aspecto relevante, considerando que atualizações dos modelos e alterações nas instruções de processamento podem modificar o comportamento de um sistema generativo. Os testes de fidelidade, segurança e equidade, nesse contexto, podem integrar procedimentos permanentes de gestão de riscos, com registros que permitam identificar modificações e avaliar sua repercussão sobre a qualidade dos resultados.

O voto do relator ressaltou essa orientação ao tratar do monitoramento da auditoria, afirmando que "O objetivo não é comprovar apenas a existência formal de controles, mas verificar sua contribuição para a redução dos riscos" (TCU, 2026, Voto, p. 11).

Referências

CONSELHO NACIONAL DE JUSTIÇA (CNJ). Resolução nº 615, de 11 de março de 2025. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025.

GOVERNMENT ACCOUNTABILITY OFFICE. **Artificial Intelligence**: An Accountability Framework for Federal Agencies and Other Entities. Washington, DC: GAO, 2021.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**. Gaithersburg: NIST, 2023. Disponível em: https://doi.org/10.6028/NIST.AI.100-1. Acesso em: 23 nov. 2026.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD AI Principles**. Paris: OECD, 2019. Disponível em: https://oecd.ai/en/ai-principles. Acesso em: 23 nov. 2026.

OWASP FOUNDATION. **OWASP Top 10 for LLM Applications 2025**. [S. l.]: OWASP Foundation, 2024. Disponível em: https://genai.owasp.org/llm-top-10/. Acesso em: 23 nov. 2026.

TRIBUNAL DE CONTAS DA UNIÃO (TCU). Acórdão nº 2.836/2026 – Plenário. Processo TC 001.458/2026-4. Relator: Ministro Odair Cunha. Sessão de 14 out. 2026. Brasília, DF: TCU, 2026.

TRIBUNAL DE CONTAS DA UNIÃO (TCU). Resolução nº 315, de 22 de abril de 2020. Dispõe sobre a elaboração de deliberações que contemplem medidas a serem tomadas pelas unidades jurisdicionadas no âmbito do Tribunal de Contas da União e revoga a Resolução-TCU 265, de 9 de dezembro de 2014. Brasília, DF: TCU, 2020.