ALUCINAÇÕES DE INTELIGÊNCIA ARTIFICIAL NO ÂMBITO JURÍDICO: ANÁLISE DA IDENTIFICAÇÃO, OCORRÊNCIA E MITIGAÇÃO DESTE FENÔMENO PELOS ASSESSORES JURÍDICOS DAS VARAS CÍVEIS DA COMARCA DE RIO BRANCO/ACRE (2023- 2025)

ARTIFICIAL INTELLIGENCE HALLUCINATIONS IN THE LEGAL SPHERE: ANALYSIS OF THE IDENTIFICATION, OCCURRENCE AND MITIGATION OF THIS PHENOMENON BY LEGAL ADVISORS OF THE CIVIL COURTS OF THE DISTRICT OF RIO BRANCO/ACRE (2023-2025)

Luis Henrique França de Oliveira[1]
Leonardo Silva de Oliveira Bandeira[2]

OLIVEIRA, Luis Henrique França; BANDEIRA, Leonardo. **Alucinações de inteligência artificial no âmbito jurídico:** análise da identificação, ocorrência e mitigação deste fenômeno pelos assessores jurídicos das varas cíveis da comarca de Rio Branco/Acre (2023-25). 28 páginas. Trabalho de Conclusão de Curso de graduação em Direito – Centro Universitário UNINORTE, Rio Branco. 2026**.**

RESUMO

Introdução. A crescente incorporação da inteligência artificial generativa no Poder Judiciário brasileiro tem promovido ganhos de eficiência e celeridade processual, mas também evidenciado riscos relacionados à geração de informações incorretas ou inexistentes, fenômeno denominado alucinação algorítmica. **Objetivo.** Analisar a identificação, ocorrência e estratégias de mitigação das alucinações de inteligência artificial adotadas por assessores jurídicos das Varas Cíveis da Comarca de Rio Branco/Acre, no período de 2023-2025. **Método.** Trata-se de pesquisa descritiva com abordagem mista (quantitativa e qualitativa), realizada por meio de questionário estruturado aplicado a assessores jurídicos, com posterior análise estatística descritiva e análise de conteúdo temática das respostas abertas. **Resultados.** Verificou-se que a maioria dos assessores utiliza ferramentas de IA para apoio na elaboração de minutas, pesquisa jurisprudencial e organização de informações processuais. Contudo, foram identificadas ocorrências de citações inexistentes, interpretações distorcidas e respostas desconectadas do contexto dos autos, evidenciando a necessidade de revisão humana obrigatória e protocolos formais de validação. **Conclusão.** Conclui-se que a inteligência artificial deve ser compreendida como instrumento auxiliar da atividade jurisdicional, sendo imprescindível a manutenção da supervisão humana qualificada, a implementação de mecanismos

[1] Discente do 8º período do curso de bacharelado em Direito pelo Centro Universitário Uninorte.
[2] Advogado, sócio do Casas, Ricardo, Bandeira & Soares Advocacia, Coordenador do Núcleo de Educação a Distância da Uninorte, Professor, especialista em Direito Civil, Docência do Ensino Superior e Legal Operations: Dados, Inteligência Artificial e Performance Jurídica pela PUC/PR.

institucionais de controle e o aprimoramento normativo para garantir segurança jurídica e confiabilidade processual.

Palavras-chave: inteligência artificial; alucinações algorítmicas; Poder Judiciário; segurança jurídica; mitigação de riscos.

ABSTRACT

Introduction. The increasing incorporation of generative artificial intelligence within the Brazilian Judiciary has promoted greater efficiency and procedural speed, while also revealing risks related to the generation of incorrect or non-existent information, known as algorithmic hallucination. **Objective.** To analyze the identification, occurrence, and mitigation strategies of artificial intelligence hallucinations adopted by legal advisors of the Civil Courts of the District of Rio Branco/Acre between 2023 and 2025. **Method.** This is a descriptive study with a mixed-method approach (quantitative and qualitative), conducted through a structured questionnaire applied to legal advisors, followed by descriptive statistical analysis and thematic content analysis. **Results.** The findings indicate that most advisors use AI tools to support drafting, legal research, and information organization. However, instances of fabricated case law, distorted legal interpretations, and contextually inconsistent responses were identified, reinforcing the need for mandatory human review and formal validation protocols. **Conclusion.** Artificial intelligence must be understood as a supportive instrument rather than a substitute for judicial activity, making it essential to maintain qualified human supervision, institutional control mechanisms, and regulatory improvements to ensure legal certainty and procedural reliability.

Keywords: artificial intelligence; algorithmic hallucinations; Judiciary; legal certainty; risk mitigation.

INTRODUÇÃO

A incorporação da inteligência artificial no sistema judiciário brasileiro representa uma das mais significativas transformações tecnológicas da contemporaneidade. Ferramentas baseadas em modelos de linguagem de grande escala passaram a auxiliar magistrados e assessores jurídicos na triagem processual, elaboração de minutas, organização de informações e pesquisa jurisprudencial, promovendo ganhos de eficiência e produtividade. Contudo, paralelamente aos avanços observados, emergem riscos estruturais associados ao fenômeno das chamadas “alucinações algorítmicas”, caracterizadas pela geração de informações factualmente incorretas ou inexistentes apresentadas com aparente coerência técnica.

O problema central que orienta esta pesquisa consiste em compreender como

as alucinações de inteligência artificial são identificadas, enfrentadas e mitigadas no cotidiano forense, especialmente no contexto das Varas Cíveis da Comarca de Rio Branco/Acre. A relevância do estudo decorre da necessidade de assegurar que a modernização tecnológica não comprometa princípios fundamentais do Estado Democrático de Direito, tais como a segurança jurídica, a fundamentação das decisões judiciais e a confiabilidade processual.

O objetivo geral consiste em analisar as práticas adotadas por assessores jurídicos para identificar e mitigar falhas decorrentes do uso de ferramentas de IA. Como objetivos específicos, busca-se: (i) examinar os fundamentos teóricos e regulatórios da aplicação da IA no Judiciário brasileiro; (ii) conceituar e classificar o fenômeno das alucinações algorítmicas; (iii) identificar riscos jurídicos associados à utilização de *outputs* imprecisos; e (iv) avaliar estratégias práticas de controle adotadas no âmbito local.

Metodologicamente, trata-se de pesquisa descritiva com abordagem mista, utilizando questionário estruturado aplicado a assessores jurídicos, com posterior análise estatística descritiva e análise de conteúdo temática. A pesquisa delimita-se ao período de 2023 a 2025, momento de intensificação da adoção de ferramentas de inteligência artificial generativa no Judiciário brasileiro.

A hipótese sustentada é a de que, embora a IA represente instrumento relevante de apoio à atividade jurisdicional, sua utilização sem protocolos formais de verificação e supervisão humana pode comprometer a qualidade decisória e a legitimidade institucional do Poder Judiciário.

1 FUNDAMENTOS TEÓRICOS DA INTELIGÊNCIA ARTIFICIAL NO DIREITO

A intersecção entre a tecnologia e o direito marca uma das transformações mais significativas do sistema judiciário nos tempos atuais. Tradicionalmente, o sistema jurídico é fundamentado na interpretação humana, na argumentação e na aplicação de precedentes. No entanto, o advento e a rápida evolução da Inteligência Artificial têm introduzido ferramentas e conceitos que não apenas otimizam tarefas, mas também levantam questões profundas sobre a natureza do trabalho jurídico e o futuro da justiça.

Essa transformação manifesta-se de forma concreta nas práticas jurídicas contemporâneas. Luz e Lima (2025, p. 102) explicam que:

> A confluência da inteligência artificial com o Direito acontece de forma direta, através da implementação de mecanismos jurídicos atrelados a softwares dotados de inteligência artificial, seja para promover maior eficiência em serviços jurídicos, como no seu uso pelos tribunais seja para catalogar, classificar e operacionalizar processos judiciais, seja na tentativa já testada em alguns países para que a máquina auxilie na tomada de decisões jurídicas em processos judiciais ou, ainda, no uso pela iniciativa privada de programas informáticos inteligentes para feitura e controle de contratos, pesquisa jurídica e predição de resultados jurídicos, entre outros usos. (Luz; Lima, 2025, p. 102).

1.1 CONCEITUAÇÃO e EVOLUÇÃO DA INTELIGÊNCIA ARTIFICIAL

A inteligência artificial pode ser definida amplamente como o campo da ciência da computação voltado a sistemas capazes de executar tarefas que, tradicionalmente, demandariam da inteligência humana, reconhecendo padrões, aprendizado, raciocínio lógico e processamento de linguagem natural. Para Jahanzaib e Tarique Anwer (2018), inteligência artificial se refere à possibilidade de a máquina emular comportamentos que reproduzam a inteligência humana. Já para Miles Brundage, Olle Häggström Peter J. Bentley e Thomas Metzinger (2018), inteligência artificial consistiria em sistemas aptos a desempenhar atividades que, quando realizadas por um indivíduo, exigiriam a utilização de suas capacidades cognitivas.

Essa busca por compreender e reproduzir a inteligência humana em um sistema automatizado levou, já na década de 1950, Alan Turing a propor o tão conhecido “Teste de Turing”, que buscava avaliar se uma máquina poderia exibir comportamento semelhante a de um ser humano (Turing, 1950). Nesse período surgiram os chamados sistemas simbólicos, que procuravam reproduzir regras lógicas de pensamento humano, dessa forma, trouxe a primeira geração de estudos no campo da inteligência artificial.

Apesar dos avanços iniciais, as limitações dos sistemas simbólicos revelaram a necessidade de novas abordagens na inteligência artificial. Como destacado por Schmidhuber (2022), o marco decisivo para a consolidação da inteligência artificial moderna ocorreu com a aplicação efetiva das técnicas de *machine learning* e *deep learning*, caracterizadas como a ciência da atribuição de crédito: "encontrar padrões nas observações que predizem as consequências das ações" (Schmidhuber, 2022, p. 1). Como destacam Alzubaidi *et al*. (2021), "um dos benefícios do *Deep Learning* é a capacidade de aprender quantidades massivas de dados", permitindo que esses algoritmos alcancem "resultados excepcionais em várias tarefas cognitivas

complexas, correspondendo ou mesmo superando aqueles fornecidos pelo desempenho humano".

Dentre os avanços mais recentes, destaca-se o modelo de linguagem generativa, que viabilizou o desenvolvimento das chamadas *Large Language Models* (*LLMs*). Como explicam Almeida, Mendonça e Filgueiras (2023), podem ser entendidas como sistemas computacionais desenvolvidos para lidar com a linguagem natural de forma semelhante ao modo como nós, humanos, a utilizamos. Esses modelos são treinados com enormes quantidades de texto e, a partir disso, conseguem prever a sequência de frases, identificar padrões de escrita e oferecer respostas que soam coerentes e adequadas ao contexto.

Baseando-se em mecanismos generativos, compostos por regras e instruções sobre a combinação de palavras, as *LLMs* tentam reproduzir o uso real da linguagem, evoluindo continuamente à medida que entram em contato com novos textos e interações. Nesse cenário, a arquitetura *Transformer,* proposta por Vaswani *et al*. (2017), representou um marco fundamental, ao substituir redes recorrentes e convolucionais por mecanismos de atenção totalmente paralelizáveis, o que permitiu maior eficiência e escalabilidade no treinamento de grandes modelos. Essa inovação abriu caminho para que as *LLMs* atuais alcançassem níveis elevados de desempenho, generalizando para múltiplas tarefas do processamento de linguagem natural.

1.2 APLICAÇÃO DA IA NO SISTEMA JUDICIÁRIO BRASILEIRO

É notório que o Sistema Judiciário Brasileiro detém um dos maiores sistemas de justiça do mundo, partindo dessa premissa, convém destacar a morosidade e o grande volume de processos que tramitam no poder judiciário dificultando assim a celeridade processual. Diante desse cenário, a aplicação da Inteligência artificial generativa surge como alternativa para modernizar e buscar formas de aumentar a eficiência produtiva.

Em 2020 o Conselho Nacional de Justiça (CNJ) criou a resolução n. 332/2020 que dispõe sobre ética, transparência e governança na produção e uso de Inteligência Artificial no Poder Judiciário, esse ato normativo regulamenta a utilização da IA no âmbito jurídico. Dentre seus primeiros artigos da resolução, o CNJ estabelece a implementação da IA pelo Poder judiciário, postulando objetivos de promover a prestação equitativa da jurisdição e o bem-estar dos jurisdicionados (CNJ,2020).

Através dessa resolução, os tribunais brasileiros passaram a investir em projetos de IA, cerca de 50% desses já se encontram em fase de testes e utilização. Dentre várias áreas de aplicabilidade da inteligência artificial no direito, destaca-se a distribuição automatizada dos processos, revisão dos pressupostos de admissibilidade como a prescrição, sugestão de minutas para cada caso concreto, verificação de improcedência de pedido liminar, entre várias outras funções (Rodas 2021).

Entre os projetos mais emblemáticos e pioneiros de aplicação da Inteligência artificial no âmbito jurídico, temos o "Victor" desenvolvido pelo Supremo Tribunal Federal (STF) com a finalidade de auxiliar na triagem de processos de repercussão geral. A ferramenta é capaz de analisar petições e identificar se elas possuem relação com temas já julgados, proporcionando significativa economia de tempo aos ministros.

Recentemente, o Superior Tribunal de Justiça (STJ) implementou uma ferramenta de inteligência artificial denominada "STJ Logos" que foi desenvolvida inteiramente pelo próprio Tribunal, possuindo o objetivo de modernizar a análise e elaboração dos conteúdos jurídicos. Essa ferramenta tem oferecido suporte diretamente aos gabinetes do Ministros tornando mais céleres e eficientes as decisões judiciais, suas principais funções contam com a geração de relatórios das decisões e análise da admissibilidade de agravos em recursos especiais (AREsps) (STJ, 2025).

O Tribunal de Justiça do Acre anunciou o lançamento da ADA (Assistente Digital Ampliada), primeira inteligência artificial generativa desenvolvida no âmbito do Judiciário acreano. A ferramenta foi desenvolvida pelo próprio tribunal e tem como finalidade apoiar o saneamento processual, a transcrição de audiências e a análise de documentos judiciais. Projetada para operar exclusivamente com dados internos da Justiça do Acre, a ADA incorpora mecanismos de anonimização de informações sensíveis, reforçando a segurança e a proteção de dados. De acordo com o presidente do TJAC, desembargador Laudivon Nogueira, a iniciativa representa a incorporação de tecnologias inovadoras para aprimorar a produtividade, acelerar os julgamentos e ampliar a eficiência do serviço prestado à população (Tribunal de Justiça do Estado do Acre, 2025).

A exemplo do *Victor* no Supremo Tribunal Federal e do *Logos* no Superior Tribunal de Justiça, a experiência do TJAC demonstra que o uso da inteligência artificial vem se consolidando como uma estratégia de modernização, com potencial de transformar rotinas processuais e apoiar magistrados em suas atividades

decisórias.

Apesar dos avanços obtidos, a aplicação da inteligência artificial no Sistema Judiciário brasileiro revela-se como um fenômeno marcado por dualidades. De um lado, estudos apontam benefícios concretos, como a redução do tempo de tramitação processual, a otimização de atividades de triagem e a melhoria na uniformização e qualidade da prestação jurisdicional. Esses resultados reforçam o potencial das ferramentas de IA para ampliar a eficiência e apoiar magistrados e servidores em tarefas repetitivas, liberando tempo para atividades analíticas de maior complexidade (FGV, 2023).

De outro, permanecem desafios significativos, especialmente relacionados à transparência dos algoritmos, ao risco de vieses, à necessidade de auditorias constantes e à preservação da autonomia decisória dos juízes. Nesse sentido, a incorporação da IA deve ser compreendida não como substitutiva da atividade jurisdicional, mas como instrumento de apoio, cuja utilização demanda responsabilidade, governança e constante reflexão crítica sobre seus impactos na justiça e na sociedade.

1.3 MARCO REGULATÓRIO E ASPECTOS ÉTICOS

A União Europeia, em resposta a esses desafios, trouxe um regulamento que exige transparência nos sistemas de IA utilizados em seus diversos setores e impõe regulamentações rigorosas para evitar a propagação de informações incorretas (*Regulation* (UE) 2024/1789). O modelo europeu estabelece uma classificação baseada em riscos, diferenciando sistemas de IA de acordo com seu potencial de dano.

Os marcos regulatórios analisados compartilham um objetivo fundamental: estabelecer um controle efetivo sobre a utilização da inteligência artificial, sobretudo no setor de justiça, reconhecendo os avanços em eficiência processual, mas destacando a necessidade de garantir que sua implementação ocorra dentro de parâmetros rigorosos para proteção dos direitos fundamentais.

No Brasil, embora não exista ainda um marco legal unificado e específico sobre inteligência artificial, já se verifica a construção de uma regulação setorial. A Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018) representou o primeiro passo relevante ao estabelecer garantias fundamentais de privacidade e segurança da informação, aplicáveis também a sistemas de IA. Contudo, no campo judicial, o

Conselho Nacional de Justiça (CNJ) vem assumindo esse protagonismo normativo.

Em 2020, a Resolução nº 332 estabeleceu os primeiros parâmetros éticos e de transparência para a utilização da IA no Poder Judiciário, afirmando, entre outros pontos, que “a utilização de modelos de Inteligência Artificial deve buscar garantir a segurança jurídica e colaborar para que o Poder Judiciário respeite a igualdade de tratamento aos casos absolutamente iguais” (Conselho Nacional de Justiça, 2020, art. 5º). Esse marco inicial inaugurou a preocupação com a utilização responsável da tecnologia na administração da justiça.

Avançando nessa trajetória, o CNJ editou a Resolução nº 615/2025, que trouxe diretrizes mais abrangentes sobre governança, requisitos de auditoria e estruturas de controle (Conselho Nacional de Justiça, 2025). A resolução instituiu o Comitê Nacional de Inteligência Artificial, composto por representantes do Judiciário, OAB, Ministério Público, Defensoria Pública e sociedade civil, com a finalidade de assegurar o uso ético e auditável da tecnologia. Além disso, estabeleceu mecanismos de rastreabilidade, monitoramento contínuo e exigência de transparência nos processos, reforçando a necessidade de segurança cibernética e de governança sólida.

Essa preocupação regulatória conecta-se diretamente ao debate ético. Conforme destacam Pinto e Ernesto (2022), é necessário que todos os envolvidos na área jurídica não abdiquem de um substrato mínimo ético essencial do processo de tomada de decisões, sob pena da redução do valor da Justiça a meros números estatísticos, que não atendem à realidade social que o Direito pretende regular. Daí a importância de princípios éticos orientadores da implementação dessa tecnologia no Judiciário.

Nesse contexto, é fundamental "investigar maneiras éticas, regulatórias e jurídicas para trazer mais transparência, supervisão e responsabilidade no manuseio dessa tecnologia inteligente na tomada de decisões, pois, na era das máquinas conscientes, a dignidade da pessoa humana deve ser sempre o objetivo de toda evolução tecnológica – e não o contrário" (Pinto; Ernesto, 2022, p. 943).

Conforme observam Tepedino e Silva (2019), "os desafios da inteligência artificial em matéria de responsabilidade civil" decorrem principalmente da autonomia dos sistemas algorítmicos, que podem ultrapassar o originalmente programado. Devido à autonomia, os robôs passam a demonstrar um comportamento emergente, surgindo situações de danos a terceiros nas quais as tradicionais estruturas delituais

quanto à responsabilidade civil passam a ter dificuldades de aplicação, uma vez que baseadas na culpa.

No ordenamento jurídico brasileiro, a ausência de legislação específica sobre responsabilidade civil por danos causados por IA faz com que se aplique o regime geral previsto no Código Civil, complementado pelas disposições do Código de Defesa do Consumidor quando aplicável. Como destacam Silva e Pereira (2025), "embora a legislação vigente forneça diretrizes gerais sobre responsabilidade civil, a ausência de normas específicas para a" inteligência artificial gera incertezas jurídicas significativas.

As Resoluções do CNJ, especialmente a 615/2025, buscam endereçar parcialmente essas questões ao estabelecer estruturas de governança e auditabilidade, exigindo que sempre haja um agente humano responsável pela decisão final. Contudo, como observa Melo (2024), a análise dos projetos legislativos em tramitação, especialmente o anteprojeto do Marco Legal da Inteligência Artificial e o Projeto de Lei 2338/2023, demonstra que a definição clara dos regimes de responsabilidade ainda demanda evolução normativa específica que contemple as peculiaridades dos sistemas inteligentes utilizados na administração da justiça.

2 ALUCINAÇÕES DA IA: IDENTIFICAÇÃO E IMPACTOS JURÍDICOS

2.1 CONCEITO E CARACTERÍSTICAS DAS ALUCINAÇÕES EM IA

A ascensão da inteligência artificial (IA) generativa, em especial os modelos de linguagem de grande escala (*LLMs*), trouxe à tona uma série de desafios e fenômenos inesperados. A alucinação da IA, um dos mais notáveis entre eles, descreve um comportamento específico e problemático desses sistemas. A definição desse fenômeno é crucial para a compreensão de seus impactos subsequentes, especialmente no campo jurídico, já que estudos indicam que as respostas geradas por *LLMs* podem ser imprecisas ou confusas em situações mais complexas, o que gera dificuldades de compreensão e insegurança para os usuários (Lima, 2023).

Conforme destacado por Hussam Alkaissi e Samy I. McFarlane em seu artigo de 2023, a própria IA, como o *ChatGPT*, oferece uma definição para a alucinação artificial:

> Alucinação artificial refere-se ao fenômeno de uma máquina, como um *chatbot*, gerar experiências sensoriais aparentemente realistas que não correspondem a nenhuma entrada do mundo real. Isso pode incluir

> alucinações visuais, auditivas ou outros tipos. Alucinações artificiais não são comuns em *chatbots*, pois eles são normalmente projetados para responder com base em regras e conjuntos de dados pré-programados, em vez de gerar novas informações. No entanto, houve casos em que sistemas avançados de IA, como modelos generativos, produziram alucinações, particularmente quando treinados com grandes quantidades de dados não supervisionados. Para superar e mitigar alucinações artificiais em *chatbots*, é importante garantir que o sistema seja devidamente treinado e testado usando um conjunto de dados diversificado e representativo. Além disso, a incorporação de métodos para monitorar e detectar alucinações, como avaliação humana ou detecção de anomalias, pode ajudar a resolver esse problema. (Alkaissi; Mcfarlane, 2023).

Essa definição nos revela que as alucinações da IA não são meras imprecisões ou erros de sintaxe, mas sim a criação de informações factualmente incorretas ou inventadas que a IA apresenta com total convicção. Portanto, embora as respostas geradas pela IA generativa possam parecer plausíveis, elas podem ser sem sentido ou incorretas (Özer 2024).

De acordo com Cavassane (2023), diversos fatores explicam a ocorrência de alucinações em modelos de linguagem. O autor destaca, citando Ji *et al.* (2024) e Kaddour *et al*. (2023), que a introdução de aleatoriedade na geração do *output*, ainda que aumente a diversidade de respostas, está diretamente associada ao crescimento no número de alucinações. Esse processo ocorre quando, em vez de selecionar o *token* mais provável, o modelo escolhe aleatoriamente entre as opções mais prováveis, o que amplia a variabilidade, mas compromete a precisão (JI *et al*., 2024 *apud* Cavassane, 2023, p. 8; Kaddour *et al.,* 2023 *apud* Cavassane, 2023, p. 21-22).

Essa perspectiva sobre as causas dessas alucinações nos conduz à necessidade de examinar mais detalhadamente como se manifestam os erros sistêmicos em IA generativa. Lemos estabelece uma definição operacional que abrange tanto as causas técnicas quanto as implicações conceituais das alucinações:

> [...] Portanto, a alucinação artificial é definida como um *output* que não parece correto, seja por problemas de aprendizagem do modelo, seja por base de dados erradas ou inconsistentes, podendo indicar inconsistências com o conjunto de dados de treinamento, com as configurações de parâmetros do modelo ou com a estrutura do modelo em si. Trata-se, de acordo com a proposta desenvolvida no item anterior, de um erro, pois é causado por princípios internos e lógicos do dispositivo. Esse erro gera falhas ou perturbações (Lemos, [2024], p. [79]).

Complementando a definição conceitual de Lemos (2024) sobre alucinações como erros sistêmicos, torna-se necessário examinar as causas específicas que originam esses fenômenos. A literatura técnica tem identificado diversos fatores contributivos, sendo a qualidade dos dados de treinamento um dos mais significativos. Estêvão e Estêvão destacam essa dimensão ao afirmar:

> O motivo mais evidente está relacionado com a qualidade dos dados usados no treino dos *LLM*. Se não existem dados sobre um determinado assunto nas fontes de informação usadas para o treino de um *LLM*, então esse *LLM* terá tendência para alucinar sobre esse assunto, eventualmente apresentando um texto coerente do ponto de vista gramatical, mas factualmente incorreto (Estêvão; Estêvão, 2023, p. 75).

Segundo Cossio (2025), as alucinações em *LLMs* podem ser classificadas em diferentes categorias, como intrínsecas, extrínsecas, de factualidade e fidelidade, além de manifestações específicas como erros factuais, inconsistências contextuais, lógicas, temporais e violações éticas.

Para isso foi elaborado uma tabela para melhor exemplificar os tipos de alucinações identificadas:

Tipo de Alucinação	Definição/Descrição	Exemplo
Intrínseca	Contradição ao próprio *input* ou contexto fornecido.	O texto diz que alguém nasceu em 1980, mas o modelo resume como 1975.
Extrínseca	Introdução de entidades ou fatos inexistentes, sem apoio no *input*.	"O Tigre Parisiense foi extinto em 1885" (espécie inexistente).
Factualidade	Erro objetivo em relação ao conhecimento real.	"Charles Lindbergh foi o primeiro a andar na Lua."
Fidelidade	Resposta plausível, mas que se desvia do *input* ou instrução.	Artigo diz que a FDA aprovou uma vacina, mas o resumo afirma que rejeitou.

Erros factuais / fabricações	Informação incorreta, enganosa ou inventada.	Inventar uma decisão judicial inexistente.
Inconsistência contextual	Contradição ou adição de informação não presente no contexto.	*Input*: “O Nilo nasce na África Central”; saída: “nasce nas montanhas da África Central”.
Inconsistência de instrução	Não seguir instruções explícitas do usuário.	Pedido de tradução para espanhol, mas resposta vem em inglês.
Inconsistência lógica	Contradições ou falhas de raciocínio interno.	Erro em cálculo matemático passo a passo.

Fonte: Adaptado de Cossio (2025).

2.2 MANIFESTAÇÕES NO CONTEXTO JURÍDICO

O Poder Judiciário brasileiro tem liderado as aplicações de inteligência artificial no âmbito da Administração Pública nacional, com 66% dos tribunais brasileiros desenvolvendo projetos de IA e 147 sistemas já registrados no Sinapses (CNJ, 2024).

Tal panorama de ampla adoção tecnológica, embora revele avanços significativos rumo à modernização da Justiça, enseja igualmente complexos desafios, sobretudo diante da ocorrência de conteúdos alucinatórios gerados pela IA, capazes de comprometer a higidez e a legitimidade da fundamentação das decisões judiciais.

O fenômeno das chamadas alucinações de inteligência artificial já alcançou o âmbito jurídico, com registros crescentes de sua manifestação em peças processuais. Desde 2023, foram documentados mais de 250 casos em tribunais ao redor do mundo, nos quais advogados ou partes utilizaram ferramentas de IA para elaborar petições e recursos contendo precedentes inexistentes, menções a doutrinas fabricadas e referências legais fictícias (Núcleo Jornalismo, 2025).

No Brasil, há alguns casos recentes que marcaram o cenário do uso da inteligência artificial no Judiciário Brasileiro. Em Santa Catarina, em fevereiro de 2025, um advogado usando *ChatGPT* citou casos fabricados e referências inexistentes na doutrina jurídica em um caso de reintegração de posse. O advogado recebeu uma

sanção de 10% do valor da causa e o caso foi submetido à OAB/SC (TJ-SC, 2025).

Outro exemplo é o caso do Tribunal de Justiça de São Paulo (TJSP, 2023), onde assessores jurídicos identificaram que a ferramenta de IA Sinapses estava gerando ementas com informações incompletas e até confeccionando trechos de jurisprudências inexistentes. De forma similar, o Tribunal Regional do Trabalho da 4ª Região (TRT-4, 2023) apontou que cerca de 12% dos resumos gerados por IA apresentavam omissões ou termos inadequados, reforçando a necessidade de transparência e supervisão.

A manifestação mais preocupante das alucinações de IA no contexto jurídico brasileiro tem sido a criação de jurisprudências completamente fictícias. Essas alucinações não se limitam a imprecisões menores ou interpretações questionáveis, mas envolvem a fabricação integral de decisões judiciais, incluindo números de processos, datas, relatores e ementas que jamais existiram nos tribunais superiores. Sobre isso, o Conselho Nacional de Justiça (CNJ) destaca:

> Assim, por exemplo, o uso do *ChatGPT* ou de ferramentas similares para elaboração de documentos jurídicos como sentenças pode trazer referências falsas a fatos ou citações de doutrina ou precedentes inexistentes, mas que simulam citações de textos elaboradas por humanos. Tal risco inclusive ocasionou incidente noticiado na imprensa envolvendo o Tribunal Regional da 1.º Região, em que sentença elaborada com o uso do *ChatGPT* fez referência a precedentes inexistentes, gerando constrangimento às partes e aos advogados, com repercussão negativa para o Poder Judiciário (CNJ, 2024).

Em casos observados, a IA tem apresentado interpretações completamente distorcidas de institutos jurídicos consolidados, criando "doutrinas" que contradizem entendimentos pacificados na jurisprudência.

Para que os sistemas de inteligência artificial sejam eficazes e suas predições possam ser explicadas, é fundamental que eles levem em conta as dinâmicas argumentativas do direito, e não apenas os aspectos empíricos, como apontado por Aletras *et al.* (2016) *apud* Maranhão *et al.* (2021, p. 171). A incorporação de técnicas de argumentação jurídica permitiria que os sistemas identifiquem os argumentos e, a partir disso, prevejam o resultado de forma mais precisa e passível de explicação para os observadores humanos.

As alucinações de inteligência artificial representam uma preocupação direta aos pilares fundamentais do sistema jurídico brasileiro, particularmente à segurança jurídica e à confiabilidade processual. Como observam Marinotti e Fachin (2023, p.

310), "os valores jurídicos intrínsecos à segurança jurídica ultrapassam a mera repetição de decisões para casos semelhantes ou idênticos", revelando que o fenômeno das alucinações compromete não apenas a consistência jurisprudencial, mas a própria estrutura de confiança que sustenta o ordenamento jurídico.

A problemática se agrava quando consideramos que "qualquer forma de facilitação nesse processo decisório gera um enviesamento por si só e contribui para a estagnação do desenvolvimento social" (Marinotti; Fachin, 2023, p. 299).

Nesse cenário, Manzato, Soares e Cugula (2025) observam que:

> A automação de decisões, facilitada pela IA, promete agilidade e eficiência. Contudo, também suscita questões críticas sobre a segurança jurídica, principalmente porque a substituição de julgamentos humanos por algoritmos pode resultar em decisões impessoais e potencialmente injustas, desafiando os princípios de equidade e justiça que fundamentam o ordenamento jurídico. E se a questão gera debates em relação às decisões do Poder Judiciário, também geram dúvidas em relação aos contratos digitais, pois a automação de decisões relacionadas a estes pode dificultar a identificação de erros ou vieses, comprometendo a confiança nas soluções tecnológicas. (Manzato; Soares; Cugula, 2025, p. 3)

Essa constatação suscita questionamentos fundamentais sobre os impactos estruturais das alucinações de IA no ordenamento jurídico pátrio. Se por um lado a automação promete eficiência, por outro, os casos já documentados evidenciam riscos concretos à integridade do sistema judicial. Torna-se, portanto, imperativo examinar como essas manifestações alucinatórias comprometem os pilares basilares da segurança jurídica e da confiabilidade processual.

O artigo 93, inciso IX, da Constituição Federal estabelece que "todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões" (Brasil, 1988). As alucinações de IA violam frontalmente esse preceito constitucional ao introduzir elementos fictícios na fundamentação, comprometendo não apenas a qualidade da decisão, mas sua própria validade jurídica.

A confiabilidade processual é um dos pilares da segurança jurídica e se conecta diretamente ao princípio da confiança legítima e da previsibilidade das decisões judiciais. No processo judicial, espera-se que os fundamentos das decisões estejam ancorados em elementos normativos, provas documentadas nos autos e precedentes devidamente reconhecidos.

Portanto, essas alucinações afetam diretamente a confiabilidade dos processos, uma vez que esses sistemas possuem a capacidade de gerar novos

conteúdos sobre a mesma perspectiva, o que compromete a consistência e a previsibilidade das decisões judiciais. Nesse sentido, como observa Holanda (2025):

> A emergência da IAG adiciona uma camada de complexidade ao debate, visto que esses sistemas possuem a capacidade de gerar novos conteúdos, levantando preocupações sobre a fidedignidade das informações, a segurança jurídica e a possibilidade de deturpações no processo decisório. O fenômeno das alucinações algorítmicas demonstra como esses sistemas podem apresentar resultados imprecisos, o que exige um controle rigoroso sobre sua aplicação no contexto judicial. Ademais, se a IA for amplamente utilizada nas decisões judiciais, serão inevitáveis julgamentos baseados em padrões éticos discutíveis, de modo a gerar distorções e injustiças irreparáveis, já que é impossível prever todos os comportamentos e interpretações da máquina. (Holanda, 2025, p. 90).

Essa perspectiva reforça que a confiabilidade processual depende não apenas da incorporação da IA como instrumento auxiliar, mas também da criação de mecanismos de controle, auditoria e supervisão humana, de modo a evitar que erros tecnológicos sejam transpostos automaticamente para o plano decisório.

2.3 ANÁLISE DE RISCOS E VULNERABILIDADES

As alucinações em sistemas de inteligência artificial no contexto jurídico são potencializadas por uma série de fatores interconectados que comprometem a confiabilidade e precisão das informações geradas.

Um dos fatores que potencializam essas alucinações pode ser causado pela própria forma como os dados são coletados e processados. Nesse sentido, Ji *et al*. (2023) explicam que a principal causa é a divergência entre a fonte e a referência:

> A principal causa de alucinação a partir de dados é a divergência entre fonte e referência. Essa divergência ocorre como um artefato da coleta heurística de dados ou está inevitavelmente contida nos dados devido à natureza de algumas tarefas de NLG. Quando um modelo é treinado com dados com essa divergência, ele pode ser incentivado a gerar texto que não é necessariamente fundamentado e não é fiel à fonte fornecida. (JI *et al*., 2023, p. 2)

Outra principal falha é a Aprendizagem de Representação Imperfeita. O codificador do modelo, que é responsável por entender e traduzir o texto de entrada em representações matemáticas, pode cometer erros. Se esse codificador aprende correlações erradas entre diferentes partes dos dados de treinamento, o modelo acaba gerando respostas que se desviam da entrada original (Parikh *et al*., 2020).

Por outro lado, vieses de exposição também contribui para a ocorrência de alucinações. Esse problema surge da discrepância entre o modo como o modelo é treinado e como ele funciona durante a geração. Enquanto o treinamento é baseado na verdade fundamental (referência real), a geração de texto em tempo real é autocondicionada, ou seja, o modelo se baseia em suas próprias previsões para gerar a próxima palavra. Essa diferença pode levar a uma progressão de erros, especialmente em sequências mais longas (Ranta *et al*., 2020).

Por fim, o viés de conhecimento paramétrico é um fator relevante. Modelos pré-treinados em grandes bases de dados tendem a memorizar informações. No entanto, em vez de se basear na entrada fornecida pelo usuário, eles podem priorizar esse conhecimento paramétrico (Longpré *et al.,* 2021). Essa preferência pode resultar na alucinação de "excesso de informações" na saída, que não estão presentes na fonte original, mas sim no conhecimento interno do modelo.

Compreendidos os fatores técnicos que levam à alucinação em modelos de IA, a discussão se expande para as implicações práticas desse fenômeno. Se a IA, como ferramenta, é cada vez mais empregada no âmbito judiciário, a ocorrência de erros factuais e informações inventadas passa de um problema de engenharia a uma séria questão de direito.

As consequências das alucinações em sistemas de inteligência artificial para a prestação jurisdicional são significativas e preocupantes. Quando informações incorretas ou fabricadas são incorporadas em peças processuais ou até mesmo em fundamentações judiciais, a própria segurança jurídica é colocada em xeque. Decisões baseadas em precedentes inexistentes ou dados distorcidos comprometem o princípio constitucional da fundamentação (art. 93, IX, da CF/88) e podem resultar na nulidade dos atos processuais.

Além disso, tais falhas repercutem diretamente na credibilidade do Poder Judiciário, ampliando a desconfiança social e abrindo espaço para litigiosidade ainda maior, já que decisões inconsistentes tendem a ser contestadas e reformadas. Como destacam Marinotti e Fachin (2023), a segurança jurídica ultrapassa a mera repetição de decisões semelhantes, sendo um valor essencial de estabilidade e confiança institucional, diretamente ameaçado pelo fenômeno das alucinações.

No plano prático, a incorporação de *outputs* algorítmicos pode levar o aumento do retrabalho processual, bem como a multiplicação de recursos e ao prolongamento das demandas, o que paradoxalmente agrava a morosidade que a inteligência artificial

propõe solucionar. Como observam Manzato, Soares e Cugula (2025), a automoção decisória, quando não acompanhada de supervisão prática, traz consigo riscos de impessoalidade e injustiça, desafiando os princípios de equidade e comprometendo a legiimidade das decisões. Dessa forma, a promessa de eficiência pode se converter em novos obstáculos à celeridade e à confiança processual.

Os riscos identificados revelam lacunas normativas relevantes quanto à responsabilidade civil e institucional pelos danos decorrentes das chamadas *alucinações* da inteligência artificial. O ordenamento jurídico brasileiro ainda carece de um marco específico que regule de forma clara a imputação de responsabilidade quando decisões judiciais sofrem influência indevida de sistemas inteligentes. Hoje, aplicam-se de modo subsidiário o Código Civil e, em alguns casos, o Código de Defesa do Consumidor, instrumentos que não contemplam as particularidades da autonomia algorítmica. Como observam Tepedino e Silva (2019), a própria lógica causal tradicional se enfraquece diante de sistemas capazes de gerar comportamentos emergentes.

Apesar dos avanços trazidos pelas Resoluções n. 332/2020 e n. 615/2025 do Conselho Nacional de Justiça, que preveem governança, rastreabilidade e supervisão humana, ainda persistem zonas cinzentas sobre a responsabilidade em caso de falhas graves. A questão central permanece: responderá o desenvolvedor, o tribunal que adota a tecnologia ou o magistrado que nela confia? Conforme apontam Silva e Pereira (2025), a ausência de diretrizes específicas gera insegurança jurídica e fragiliza a confiança no sistema. Nesse contexto, mostra-se urgente o avanço de propostas legislativas, como o Projeto de Lei n. 2.338/2023 e o anteprojeto do Marco Legal da Inteligência Artificial, que possam estabelecer parâmetros claros e proporcionais aos riscos.

Em síntese, as falhas da IA não se limitam ao campo técnico: repercutem na própria legitimidade da justiça. Por isso, impõe-se a criação de mecanismos de controle contínuo e de normas que assegurem transparência, responsabilização e proteção de direitos fundamentais. Sem tais garantias, o potencial modernizador da IA no Judiciário corre o risco de ser ofuscado por efeitos adversos que fragilizam a função jurisdicional.

3 ESTRATÉGIAS DE MITIGAÇÃO: ANÁLISE EMPÍRICA NA COMARCA DE RIO BRANCO/AC

3.1 METODOLOGIA E CARACTERIZAÇÃO DA PESQUISA

Este estudo constitui-se, quanto ao objetivo, como uma pesquisa descritiva com tratamento misto dos dados (quantitativo e qualitativo). A investigação teve por finalidade mapear as percepções, práticas e estratégias de mitigação relacionadas às chamadas “alucinações” produzidas por sistemas de Inteligência Artificial entre assessores jurídicos das Varas Cíveis da Comarca de Rio Branco (AC) e caracterizar procedimentos de validação adotados no ambiente de trabalho. A escolha pela abordagem mista justifica-se pela necessidade de quantificar padrões de uso e percepção (variáveis fechadas) e, simultaneamente, compreender com profundidade incidentes, exemplos e sugestões trazidas pelos participantes (respostas abertas).

População e amostra. A população alvo do estudo foi composta por assessores jurídicos que atuam em varas cíveis e por assessores que atuam nas câmaras cíveis da Comarca de Rio Branco/AC. A amostra é constituída pelos respondentes do questionário aplicado por meio da plataforma *Google Forms;* o instrumento registrou as respostas de todos os participantes que consentiram com a participação anônima.

Instrumento de coleta. O instrumento utilizado foi um questionário estruturado hospedado em *Google Forms,* composto por sete seções: (A) informações sociodemográficas e profissionais (código do participante, faixa etária, gênero, tempo de atuação, nível de formação e unidade de atuação); (B) uso de ferramentas de IA (frequência, tipos de ferramenta e finalidades); (C) experiências com respostas incorretas ou “alucinações” geradas por IA (ocorrência, frequência, descrição de casos e gravidade percebida); (D) métodos de detecção e validação utilizados no cotidiano; (E) formação e percepção sobre preparo para identificação de falhas de IA; (F) políticas institucionais e sugestões de protocolos; e (G) considerações finais e autorização de uso anônimo dos dados.

Procedimento de aplicação. O questionário foi enviado eletronicamente aos assessores por meio de *link* compartilhado internamente. As respostas foram coletadas de forma voluntária e anônima; o campo de identificação do respondente possui apenas um código (iniciais) para permitir identificação interna sem expor dados pessoais. Tratamento dos dados e técnicas de análise: Análise quantitativa (dados fechados): foram calculadas frequências absolutas e relativas (percentuais) para todas as variáveis categóricas (por exemplo: frequência de uso de IA; tipos de ferramentas

utilizadas; finalidades principais de uso; existência de procedimentos formais; escala de preparo para identificação de falhas). Análise qualitativa (respostas abertas): as respostas textuais foram submetidas à análise de conteúdo temática, visando identificar categorias recorrentes relacionadas a: (i) tipos de alucinações observadas (ex.: jurisprudência inventada; citação de norma inexistente; falta de relação entre resposta e processo anexado); (ii) impactos percebidos (insignificante, baixa, média, alta, crítica); (iii) estratégias e protocolos sugeridos (registro do uso, revisão humana obrigatória, lista de ferramentas autorizadas, entre outros); e (iv) exemplos e narrativas que ilustrem casos representativos. O procedimento de codificação seguirá etapas de leitura flutuante, categorização inicial, refinamento das categorias e quantificação de menções quando relevante, permitindo integrar os achados qualitativos com os dados quantitativos. Essa combinação (triangulação de métodos) visa conferir maior robustez interpretativa aos resultados descritos neste capítulo.

Limitações metodológicas. Entre as limitações reconhecidas destacam-se: (a) a natureza não probabilística da amostra (autorreporte voluntário) que impede generalizações amplas para além do contexto da Comarca estudada; (b) o tamanho amostral (n = 10) — suficiente para análise descritiva local, mas com limites para inferências mais amplas; (c) possíveis vieses de desejabilidade social nas respostas (respondentes podem subestimar práticas de não verificação); e (d) heterogeneidade na interpretação de termos técnicos (por exemplo, “alucinação”), o que motivou a inclusão de questão aberta para captura de exemplos concretos.

3.2 PERCEPÇÕES E PRÁTICAS DOS ASSESSORES JURÍDICOS

A presente seção apresenta a análise integrada (quantitativa-descritiva e qualitativa-interpretativa) dos dados coletados junto a assessores jurídicos das Varas Cíveis da Comarca de Rio Branco/AC, com foco nas práticas de utilização de ferramentas de Inteligência Artificial (IA), nas experiências relativas a respostas incorretas ou alucinações geradas por IA, nos métodos de detecção/validação adotados e na relação entre formação profissional e procedimentos de mitigação. As interpretações seguem a lógica da triangulação entre as respostas fechadas e os relatos abertos fornecidos pelos participantes, conforme previsto no delineamento metodológico deste estudo.

3.2.1 Uso de ferramentas de IA: finalidade e frequência.

Os dados coletados indicam que o uso de ferramentas baseadas em IA já está incorporado à rotina de trabalho de boa parte dos assessores, sendo registradas múltiplas finalidades práticas. Entre as finalidades mais recorrentes destacam-se: redação de minutas/peças, pesquisa jurisprudencial, elaboração de resumos de processos e geração de rascunhos/ideias iniciais. As ferramentas mencionadas pelos respondentes incluem *chatbots* de linguagem (ex.: *ChatGPT*), ferramentas de pesquisa jurisprudencial com IA e ferramentas de resumo automático, o que demonstra um emprego instrumental da tecnologia para tarefas de apoio à produção textual e à busca de precedentes.

Interpreta-se esse padrão como reflexo de uma adoção pragmática: os assessores tendem a utilizar a IA como aceleradora de tarefas repetitivas e como "assistente de ideação", preservando a atividade decisória e de validação final para o trabalhador humano. Todavia, o uso difundido aumenta a exposição a resultados automatizados potencialmente imprecisos, condição que torna imprescindível a implementação de rotinas de checagem e de protocolos institucionais que mitiguem riscos processuais e reputacionais.

3.2.2 Ocorrência e tipologia das "alucinações" identificadas

Os relatos abertos dos participantes apontam três tipos de erro gerados por IA que se repetem com frequência: Jurisprudência inventada ou citações jurisprudenciais imprecisas. Vários respondentes relataram que a ferramenta sugere e cita acórdãos ou ementas que não existem, ou atribui trechos a processos diferentes; Citação de normas inexistentes. Foram registradas menções a artigos/leis/normas que, conforme verificação posterior, não constam na legislação; Geração de minutas com dados de outro processo ou sem relação com os documentos anexados. Houve relatos de produção de texto irrelevante ao caso concreto, incluindo uso indevido de informações de processos distintos.

Do ponto de vista da gravidade, os respondentes variaram na avaliação do impacto, desde efeitos insignificantes (erros facilmente detectáveis em revisão) até impactos classificados como "médios/altos" quando a informação incorreta passou despercebida e chegou a ser incorporada em documentos oficiais. A presença de relatos que descrevem casos em que a IA produziu minutas incompatíveis com os autos ou jurisprudência inventada sugere risco real de prejuízo prático quando não há

rotina rigorosa de validação humana.

3.2.3 Procedimentos de detecção e validação adotados

Os métodos de detecção informados pelos assessores mostram uma combinação de procedimentos formais e informais, com destaque para: verificação de fontes (checagem das referências e do texto gerado contra bases oficiais e doutrina); cruzamento com jurisprudência em bases reconhecidas; consulta às bases oficiais (portais dos tribunais, sistemas oficiais de legislação); revisão por colega ou supervisor como mecanismo de controle adicional; intuição e experiência profissional para identificar respostas com sinais de incoerência ou imprecisão.

Esses procedimentos indicam que, na prática, a validação é predominantemente baseada em verificação humana, apoiada por consulta a fontes oficiais, ou seja, uma resposta coerente ao risco específico das “alucinações” (erros factuais e citações fictícias). No entanto, os dados também revelam que nem todos os gabinetes dispõem de procedimentos formais padronizados: algumas unidades relataram existir apenas práticas informais ou nenhuma diretriz institucional para validação, o que aumenta a variabilidade na qualidade do controle interno e potencialmente eleva a probabilidade de erro sistêmico quando a IA é usada de maneira não acompanhada.

3.2.4 Relação entre formação/experiência e práticas de verificação

A análise qualitativa das respostas sugere que o nível de formação e o tempo de atuação influenciam as práticas de uso e checagem. De modo geral: profissionais com maior experiência prática frequentemente relatam maior habilidade para identificar incoerências geradas por IA com base em “intuição profissional” e hábito de cruzamento de fontes; respondentes com formação de pós-graduação ou mestrado tendem a valorizar a verificação documental e a indicar a necessidade de capacitação específica sobre verificações técnicas de *outputs* gerados por IA; contudo, houve respostas que apontam lacuna na formação formal sobre o uso seguro de IA: muitos manifestaram não ter recebido treinamento estruturado, o que evidencia uma demanda por capacitação prática e por protocolos institucionais formais de checagem.

Interpretativamente, esses achados indicam que experiência e formação ajudam na detecção intuitiva de erros, mas não substituem a necessidade de rotinas

formais e de treinamentos que tornem a prática de verificação consistente entre diferentes gabinetes e níveis de senioridade.

3.3 PROPOSIÇÕES PARA USO SEGURO DA IA PERCEPÇÃO DE RISCOS E PROPOSTAS DE PROTOCOLOS INSTITUCIONAIS

Entre os riscos apontados pelos participantes, salientam-se: decisões automatizadas sem análise do caso concreto, incorporação de jurisprudência falsa em peças processuais, exposição de dados sensíveis e redundância de trabalho quando a IA gera versões inexatas que demandam retrabalho. Como protocolos sugeridos para mitigar essas falhas, os respondentes propuseram recorrência em itens semelhantes: revisão humana obrigatória de qualquer conteúdo jurídico produzido pela IA; registro do uso da IA (logs de quem usou, para qual finalidade e qual ferramenta); treinamento periódico sobre práticas seguras de uso de IA e sobre identificação de sinais de alucinação; lista de ferramentas autorizadas e padronizadas pelo tribunal/instituição; política de não uso em matérias sensíveis que possam envolver segredo de justiça, dados pessoais críticos ou decisões de alta relevância; procedimentos formais de checagem documental, incluindo cruzamento obrigatório com bases oficiais antes de anexação de conteúdo ao processo.

Essas sugestões representam um repertório prático que pode ser operacionalizado em políticas institucionais e em protocolos de gabinetes, reduzindo a assimetria de práticas entre assessores e mitigando riscos processuais decorrentes de saídas automatizadas imprecisas.

3.3.1 Interpretação crítica e implicações para mitigação

A interpretação integrada dos dados aponta para uma situação ambivalente: a IA é percebida como instrumento capaz de aumentar a eficiência (sobretudo em tarefas rotineiras), mas essa vantagem só se materializa se houver supervisão humana eficaz e procedimentos formais de validação. Em termos práticos, o estudo indica três linhas prioritárias de intervenção institucional: formalização de procedimentos de validação: padronizar que qualquer saída da IA utilizada em peças jurídicas passe por verificação documental por servidor ou assessor responsável. Capacitação dirigida: oferecer treinamentos práticos (workshops com simulações de

alucinações, uso de bases oficiais, técnicas de *prompt* e de verificação) que articulem conhecimentos técnicos e legais. Governança do uso de ferramentas: adotar políticas internas que definam ferramentas autorizadas, critérios de registro do uso e restrições para matérias sensíveis.

Essas medidas alinham-se com boas práticas de governança de tecnologia e com o princípio de que a automação deve ser complementada por controles humanos e institucionais para preservar a segurança jurídica e a integridade do processo decisório.

3.3.2 Considerações interpretativas e implicações práticas

A análise integrada dos resultados permite afirmar que os assessores jurídicos da Comarca de Rio Branco/AC demonstram postura predominantemente prudente e reflexiva frente à IA: reconhecem seu potencial de otimização, mas também os riscos inerentes à automatização da linguagem jurídica.

De forma sintética, pode-se concluir que: A IA já é amplamente utilizada como ferramenta auxiliar de produtividade; há conhecimento empírico sobre a ocorrência de alucinações e métodos de verificação, embora ainda sem padronização institucional; E há consenso quanto à necessidade de políticas internas e capacitações contínuas.

A adoção da IA, nesse contexto, representa não apenas um avanço técnico, mas um desafio ético e organizacional. A mitigação eficaz de alucinações exige integração entre controle humano, capacitação contínua e diretrizes institucionais, assegurando que o uso da tecnologia reforce, e não fragilize a segurança e a legitimidade das decisões judiciais.

3.3.3 Diretrizes para implementação responsável da IA no ambiente jurídico

Com base nas boas práticas internacionais e nos resultados desta pesquisa, recomenda-se a adoção de diretrizes institucionais para orientar o uso responsável da IA no Poder Judiciário. Essas diretrizes devem se apoiar em quatro princípios norteadores: transparência e rastreabilidade: todos os documentos produzidos com apoio de IA devem conter registro claro de sua origem e das etapas de validação realizadas, de forma auditável. Proporcionalidade e finalidade legítima: a utilização da IA deve sempre ter propósito jurídico ou administrativo definido, evitando usos recreativos, indevidos ou fora do escopo da função pública. Segurança da informação: ferramentas de IA só devem ser empregadas quando garantirem proteção de dados

sigilosos e conformidade com a Lei Geral de Proteção de Dados (Lei nº 13.709/2018). Responsabilidade compartilhada: a responsabilidade pelo uso da IA é institucional, cabendo aos órgãos gestores estabelecer normas e sistemas de controle, e aos usuários, observar as boas práticas operacionais.

Essas diretrizes favorecem a construção de uma governança tecnológica madura**,** capaz de equilibrar inovação e segurança, estimulando o uso produtivo da IA sem comprometer a integridade jurídica das decisões e atos processuais.

3.3.4 Síntese interpretativa

A partir das evidências empíricas e da análise interpretativa dos dados, observa-se que o uso da IA no contexto jurídico é inevitável e, ao mesmo tempo, requer limites claros. As proposições apresentadas neste item configuram um modelo preventivo de adoção tecnológica, centrado no controle humano e na institucionalização de rotinas seguras.

A efetividade dessas medidas depende da integração entre conhecimento técnico, ética profissional e estrutura normativa, de modo que o uso de IA se torne uma extensão racional e controlada da capacidade humana, e não um substituto da reflexão jurídica.

CONSIDERAÇÕES FINAIS

A presente pesquisa evidenciou que a utilização da inteligência artificial no âmbito do Poder Judiciário brasileiro constitui realidade irreversível e em constante expansão. No contexto específico da Comarca de Rio Branco/Acre, verificou-se que os assessores jurídicos já incorporaram ferramentas de IA em suas rotinas, sobretudo para elaboração de minutas, síntese de informações processuais e apoio à pesquisa jurisprudencial.

Entretanto, os dados empíricos demonstraram que o fenômeno das alucinações algorítmicas não é meramente hipotético, mas uma ocorrência concreta e identificável. Foram relatados casos de citações jurisprudenciais inexistentes, interpretações equivocadas de dispositivos legais e respostas desconectadas do contexto processual analisado. Tais ocorrências reforçam que a inteligência artificial generativa, embora sofisticada, não possui compreensão semântica real, operando com base em probabilidades estatísticas.

Conclui-se que a mitigação desses riscos depende fundamentalmente da manutenção da supervisão humana qualificada, da implementação de protocolos institucionais de validação obrigatória e da capacitação contínua dos profissionais que utilizam tais ferramentas. A IA deve ser compreendida como instrumento auxiliar e não como substituta da atividade jurisdicional, preservando-se a responsabilidade decisória do magistrado e a análise crítica dos assessores.

Além disso, destaca-se a necessidade de avanço normativo específico acerca da responsabilidade civil decorrente de falhas algorítmicas, bem como o fortalecimento das diretrizes de governança previstas nas Resoluções do Conselho Nacional de Justiça. A consolidação de uma cultura de uso responsável e auditável da inteligência artificial revela-se condição essencial para que a modernização tecnológica caminhe em consonância com os princípios constitucionais da segurança jurídica, transparência e fundamentação das decisões judiciais.

Por fim, entende-se que o debate sobre alucinações de IA no Direito transcende a dimensão técnica e insere-se no campo ético-institucional, exigindo permanente reflexão crítica sobre os limites e possibilidades da tecnologia na promoção da justiça.

REFERÊNCIAS

ALETRAS, Nikolaos *et al*. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. **PeerJ Computer Science**, [s. l.], v. 2, n. e93, 24 out. 2016. DOI: 10.7717/peerj-cs.93. Disponível em: https://peerj.com/articles/cs-93/. Acesso em: 5 set. 2025.

ALMEIDA, V.; MENDONÇA, R. F.; FILGUEIRAS, F. *ChatGPT:* tecnologia, limitações e impactos. Ciência Hoje, Rio de Janeiro, p. 1–11, 2024. Disponível em: https://cienciahoje.org.br/artigo/*chatgpt*-tecnologia-limitacoes-e-impactos/#. Acesso em: 21 abr. 2025.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 332, de 21 de agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 24 ago. 2025.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 615, de 2025**. Regulamenta o uso da inteligência artificial no Poder Judiciário. Portal CNJ, 2025. Disponível em: https://hdl.handle.net/20.500.12178/247151 Acesso em: 24 ago. 2025.

BRASIL. **Constituição da República Federativa do Brasil de 1988.** Brasília, DF:

Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 5 set. 2025.

BRASIL. Superior Tribunal de Justiça (STJ). **STJ lança novo motor de inteligência artificial generativa para aumentar eficiência na produção de decisões**. Notícias, 11 fev. 2025. Disponível em: https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2025/11022025STJ-lanca-novo-motor-de-inteligencia-artificial-generativa-para-aumentar-eficienciana-producao-de-decisoes.aspx. Acesso em: 24 ago. 2025.

BRUNDAGE, Miles; HÄGGSTRÖM, Olle; BENTLEY, Peter J.; METZINGER, Thomas. Should we fear artificial intelligence? Scaling Up Humanity: The Case for Conditional Optmism about Artificil Intelligene. **EPRS. European Parlimentary Ressearch Service.** Bruxelas, mar 2018. Disponível em https://europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf. Acesso em: 23 ago. 2025.

CAVASSANE, Ricardo Peraça. **Verdade e falsidade no *ChatGPT:*** uma análise de *chatbots* baseados em *Large Language Models* da perspectiva da teoria semântica da verdade de Tarski. Campinas: Unicamp/CLE, 2023. Acesso em: 31 ago. 2025.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução n. 332 de 21 de agosto de 2020**. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em 24 ago. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **O uso da Inteligência Artificial Generativa no Poder Judiciário Brasileiro**. Brasília, DF: CNJ, 2024. Relatório técnico. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2024/09/cnj-relatorio-depesquisa-iag-pj.pdf. Acesso em: 5. set. 2025.

COSSIO, Manuel. ***A comprehensive taxonomy of hallucinations in Large Language Models*.** Universitat de Barcelona, 2025. Disponível em: https://arxiv.org/abs/2508.01781. Acesso em: 2 set. 2025.

ESTÊVÃO, João M. C.; ESTÊVÃO, M. Dulce. **Inteligência Artificial na avaliação tradicional:** aquisição de conhecimento vs *Prompt* Engineering. In: CONGRESSO NACIONAL DE PRÁTICAS PEDAGÓGICAS NO ENSINO SUPERIOR, 9., 2023, Faro. Livro de Atas... Faro: Universidade do Algarve, 2023. p. 73-289. Acesso em: 2 ago. 2025.

FUNDAÇÃO GETULIO VARGAS (FGV). Relatório Inteligência Artificial no Judiciário Brasileiro – *3ª edição*. São Paulo: FGV, 2023. Acesso em: 6 set. 2025.

JI, Ziwei *et al*. *Survey of Hallucination in Natural Language Generation*. **ACM Comput. Surv.**, v. 55, n. 12, p. 248, dez. 2023. Disponível em: https://dl.acm.org/doi/10.1145/3571730. Acesso em: 7 set. 2025.

LEMOS, André Luiz Martins. **Erros, falhas e perturbações digitais em alucinações das**
IA generativas: tipologia, premissas e epistemologia da comunicação. Matrizes, São Paulo, v. 18, n. 1, p. 75-91, 2024. DOI: https://doi.org/10.11606/issn.1982-8160.v18i1p75-91. Disponível em: https://revistas.usp.br/matrizes/article/view/210892. Acesso em: 2 set. 2025.

LIMA, J. Como o *chatgpt* afeta a educação e o desenvolvimento universitário. The trends hub, n. 3, 2023.
LONGPRÉ, Shayne *et al*. Entity-Based Knowledge Conflicts in Question Answering. **arXiv**, 2021. Disponível em: https://arxiv.org/abs/2109.05052. Acesso em: 7 set. 2025.

LUZ, Rodrigo Rodrigues da; LIMA, Marília Freitas. Inteligência artificial e direito: um estudo sobre os impactos nas dimensões dos direitos humanos fundamentais. **Revista de Direito, Governança e Novas Tecnologias**, v. 10, n. 2, p. 101-123, jan./jul. 2025. Acesso em: 7 set. 2025.

MARANHÃO, Juliano Souza de Albuquerque; FLORÊNCIO, Juliana Abrusio; ALMADA, Marco. Inteligência artificial aplicada ao direito e o direito da inteligência artificial. **Suprema – Revista de Estudos Constitucionais**, Distrito Federal, Brasil, v. 1, n. 1, p. 154–180, 2021. DOI: 10.53798/suprema.2021.v1.n1.a20. Disponível em: https://suprema.stf.jus.br/index.php/suprema/article/view/20. Acesso em: 5 set. 2025.

MELO, Guilherme da Silva. **Inteligência Artificial e responsabilidade civil: uma análise do anteprojeto do Marco Legal da Inteligência Artificial e do Projeto de Lei 2338/2023**. Revista IBERC, Belo Horizonte, v. 7, n. 1, p. 49-65, 2024. Acesso em: 25 set. 2025.

NÚCLEO JORNALISMO. **Alucinações de IA em processos jurídicos**. 2025. Disponível em: https://nucleo.jor.br/interativos/2025-08-07-alucinacoes-ia-processos-juridicos/. Acesso em: 5 set. 2025.

ÖZER, M. Is *Artificial Intelligence hallucinating?* **Turk Psikiyatri Derg**, v. 35, n. 4, p. 333-335, 14 out. 2024. DOI: 10.5080/u27587. Disponível em: https://pubmed.ncbi.nlm.nih.gov/39398861/. Acesso em: 30 ago. 2025.

PARIKH, Ankur P. *et al.* ToTTo: **A Controlled Table-To-Text Generation *Dataset***. In: Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020. Disponível em: https://aclanthology.org/2020.emnlp-main.89. Acesso em: 7 set. 2025.

PINTO, Henrique Alves; ERNESTO, Leandro Miranda. **Inteligência Artificial aplicada ao Direito: por uma questão de ética.** Revista de Jurisprudência do CIDP, v. 6, p. 919-946, 2022. Disponível em: https://www.cidp.pt/revistas/rjlb/2022/6/2022_06_0919_0946.pdf. Acesso em: 30 ago. 2025.

RANTA, Marc'Aurelio *et al*. Sequence Level be with Recurrent Neural Networks. In: **ICLR 2016: 4th International Conference on Learning Representations**. San Juan,

Porto Rico, 2016. Disponível em: https://arxiv.org/abs/1511.06732. Acesso em: 7 set. 2025.

RODAS, Sérgio. **Metade dos tribunais brasileiros já tem sistemas de inteligência artificial.**In: Conjur [on-line].Disponível em: https://www.conjur.com.br/2021-jul11/metade-cortes-brasileiras-projeto-inteligencia-artificial. Acesso em: 24 ago. 2025.

SHABBIR, J., & ANWER, T. (2018***). Artificial Intelligence and its Role in Near Future***. *arXiv: Artificial Intelligence*. https://arxiv.org/pdf/1804.01396.pdf. Acesso em: 23 ago. 2025.

SILVA, Marina Arista; PEREIRA, Camilly Vitoria das Chagas. **A responsabilidade civil no uso de Inteligência Artificial**. Migalhas, jan. 2025. Disponível em: https://www.migalhas.com.br/depeso/422333/a-responsabilidade-civil-no-uso-deinteligencia-artificial. Acesso em: 24 ago. 2025.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil,** Belo Horizonte, v. 21, n. 3, p. 61-85, 2019. Acesso em: 28 set. 2025.

TRIBUNAL DE JUSTIÇA DE SANTA CATARINA (TJSC). *TJSC* **multa autor de recurso por jurisprudência falsa gerada por IA**. Disponível em: https://www.tjsc.jus.br/web/imprensa/-/tjsc-multa-autor-de-recurso-por-jurisprudenciafalsa-gerada-por-ia?ref=nucleo.jor.br. Acesso em: 5 set. 2025.

TRIBUNAL DE JUSTIÇA DO ACRE (TJAC). **ADA***:* primeira inteligência generativa criada pelo TJAC será lançada na terça-feira, 12. Notícias, 24 ago. 2025. Disponível em: https://www.tjac.jus.br/2025/08/ada-primeira-inteligencia-generativa-criada-pelotjac-sera-lancada-na-terca-feira-12/. Acesso em: 24 ago. 2025.

TRIBUNAL REGIONAL DO TRABALHO DA 4a REGIÃO. **Relatório de impacto da IA na elaboração de acórdãos**. Porto Alegre, 2023. Disponível em: https://www.trt4.jus.br/. Acesso em: 30 mar. 2025.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1789 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, relativo à inteligência artificial (Lei de IA da UE).** Bruxelas, 2024. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 30 mar. 2025.

VASWANI, A. *et al*. ***Attention is all you need***. *Advances in neural information processing systems*, v. 30, 2017. Acesso em: 30 julho. 2025.

EDITORA

MOMENTO JURÍDICO