Editora responsável: Profa. Dra. Fayga Bedê
https://orcid.org/0000-0001-6444-2631

doi:10.12662/2447-6641oj.v23i44.p1-26.2025

Submetido: 30 jun. 2025
Aprovado: 19 set. 2025

# Entre Alucinações e Inovações: o Desafio da Implementação Responsável da IA Generativa no Sistema de Justiça

# Between Hallucinations and Innovations: the Challenge of Responsible Implementation of Generative AI in the Justice System

# Entre Alucinaciones e Innovaciones: el Desafío de la Implementación Responsable de la IA Generativa en el Sistema de Justicia

Cíntia Menezes Brunetta[*]
Taís Schilling Ferraz[**]
Alisson Carvalho de Alencar[***]



[*] Doutora em Direito pela Faculdade Autônoma de Direito (Fadisp). Conselheira do Conselho Nacional do Ministério Público (CNMP). Juíza federal do Tribunal Regional Federal da 5 Região. Professora da Graduação e dos Programas de Mestrado e Doutorado da Fadisp. Pesquisadora da Fundação Nacional de Desenvolvimento do Ensino Superior Particular (Funadesp). Coordenadora Adjunta do Mestrado Interinstitucional Fadisp/Ejug. Formadora de magistrados e formadora de formadores da Enfam e de outras escolas de magistratura e do Ministério Público. Fortaleza - CE - BR. E-mail: cbrunetta@hotmail.com. Orcid: https://orcid.org/0000-0001-5610-1693

[**] Doutora em Ciências Criminais e Mestre em Direito pela PUCRS. Desembargadora do TRF4. Professora do quadro permanente e vice-coordenadora do Programa de amestrado da ENFAM e membro do Centro Nacional de Inteligência da JF. Brasília - DF - BR. E-mail: taissferraz@gmail.com. https://orcid.org/0000-0002-9592-0488

[***] Pós-doutorando pela Universidade de São Paulo (USP); Doutor em Direito pela Universidade de Salamanca, em convênio com a Faculdade Autônoma de Direito (2020); Mestre em Administração Pública pela Escola Brasileira de Administração Pública da Fundação Getúlio Vargas (2015). Procurador Geral de Contas do Ministério Público de Contas do Estado de Mato Grosso. Cuiabá - MT - BR. E-mail: alisson.alencar@unialfa.com.br. https://orcid.org/0009-0007-6614-5025


Esta obra está licenciada sob uma licença Creative Commons Atribuição - Não comercial - Compartilhar igual 4.0 Internacional.

## RESUMO

**Contexto**: As potencialidades da Inteligência Artificial Generativa – IAG no Poder Judiciário, a par de produzirem grandes expectativas quanto ao aumento da eficiência, têm motivado grandes preocupações, reproduzindo as perplexidades que sempre permearam as relações da humanidade com a inovação, notadamente a tecnológica.

**Objetivos**: A investigação que originou este artigo teve dois objetivos: reunir e sistematizar alguns conceitos e conhecimentos sobre a IAG, suas características, sua evolução e seus modelos de linguagem; e avaliar de que maneira a IAG pode ser empregada no contexto do Judiciário, com adequado gerenciamento de riscos e aproveitamento de potencialidades. A hipótese a ser explorada é de que os riscos da Inteligência Artificial Generativa (IAG) no sistema de justiça não estão apenas em seu uso inadequado, mas também em sua não utilização.

**Método**: A pesquisa é bibliográfica e documental. Parte-se de um histórico sobre a computação e a inteligência artificial - IA, apresentam-se noções gerais sobre IA, Machine Learning e IA Generativa, explorando-se, na sequência, os riscos do desconhecimento das características e das limitações da IAG. Explora-se, então, como está sendo absorvida essa inovação no Poder Judiciário, identificando-se seus potenciais benefícios.

**Resultados**: Os resultados revelam que há riscos de alucinações nas respostas, na reprodução de vieses e no manejo inadequado de dados sensíveis, bem como que é possível gerenciá-los, mediante adequado conhecimento sobre o funcionamento e as limitações da IAG, governança pautada na ética e no protagonismo humano. Revelam, também, que, diante do volume de dados e informações que transitam e são gerados pelo Poder Judiciário, o uso da IAG vem se tornando condição de possibilidade na gestão do conhecimento para uma adequada e coerente prestação jurisdicional.

**Conclusões**: Concluiu-se pela confirmação da hipótese, identificando-se que os riscos da IAG no Judiciário residem tanto em não saber utilizá-la como também em não a utilizar.

**Palavras-chave:** Inteligência Artificial Generativa; riscos e potencialidades; Poder Judiciário; protagonismo humano; governança.

## ABSTRACT

**Context**: The potential of Generative Artificial Intelligence (GAI) in the Judiciary, while generating great expectations regarding increased efficiency, has motivated significant concerns, reproducing the perplexities that have always permeated humanity's relationship with innovation, notably technological innovation.

**Goal**: The investigation that gave rise to this article had two objectives: to gather and systematize some concepts and knowledge about GAI, its characteristics, evolution, and language models; and to assess how GAI can be employed in the context of the Judiciary,

with proper risk management and harnessing of its potential. The hypothesis to be explored is that the risks of Generative Artificial Intelligence (GAI) in the justice system lie not only in its misuse but also in its non-use.

**Method**: The research is bibliographic and documentary. It begins with a historical overview of computing and artificial intelligence – AI, presents general notions about AI, Machine Learning, and Generative AI, then explores the risks of ignoring the characteristics and limitations of GAI. It then examines how this innovation is being absorbed by the Judiciary, identifying its potential benefits.

**Results**: The results reveal that there are risks of hallucinations in responses, reproduction of biases, and inadequate handling of sensitive data, but also that these risks can be managed through proper knowledge of GAI's functioning and limitations, governance guided by ethics, and human protagonism. They also show that, given the volume of data and information that circulate and are generated by the Judiciary, the use of GAI is becoming a condition of possibility in knowledge management for an adequate and coherent judicial performance.

**Conclusions**: The hypothesis was confirmed, identifying that the risks of GAI in the Judiciary lie both in not knowing how to use it and in not using it at al.

**Keywords:** Generative Artificial Intelligence; risks and potential; Judiciary; human protagonism; governance.

**RESUMEN**

**Contextualización**: Las potencialidades de la Inteligencia Artificial Generativa (IAG) en el Poder Judicial, además de producir grandes expectativas en cuanto al aumento de la eficiencia, han motivado grandes preocupaciones, reproduciendo las perplejidades que siempre han permeado las relaciones de la humanidad con la innovación, notablemente la tecnológica.

**Objectivos**: La investigación que dio origen a este artículo tuvo dos objetivos: reunir y sistematizar algunos conceptos y conocimientos sobre la IAG, sus características, evolución y modelos de lenguaje; y evaluar de qué manera la IAG puede ser empleada en el contexto del Poder Judicial, con una adecuada gestión de riesgos y aprovechamiento de sus potencialidades. La hipótesis a explorar es que los riesgos de la Inteligencia Artificial Generativa (IAG) en el sistema de justicia no se encuentran solo en su uso indebido, sino también en su no utilización.

**Método**: La investigación es bibliográfica y documental. Parte de un recorrido histórico sobre la computación y la inteligencia artificial – IA, presenta nociones generales sobre IA, Machine Learning e IA Generativa, y luego explora los riesgos de desconocer las características y limitaciones de la IAG. Posteriormente, se analiza cómo esta innovación

está siendo absorbida por el Poder Judicial, identificando sus potenciales beneficios.

**Resultados**: Los resultados revelan que existen riesgos de alucinaciones en las respuestas, reproducción de sesgos y manejo inadecuado de datos sensibles, pero también que es posible gestionarlos mediante un adecuado conocimiento sobre el funcionamiento y las limitaciones de la IAG, una gobernanza guiada por la ética y el protagonismo humano. Asimismo, muestran que, dado el volumen de datos e informaciones que circulan y son generados por el Poder Judicial, el uso de la IAG se está convirtiendo en una condición de posibilidad en la gestión del conocimiento para una prestación jurisdiccional adecuada y coherente.

**Conclusiones**: Se confirmó la hipótesis, identificándose que los riesgos de la IAG en el Poder Judicial residen tanto en no saber utilizarla como en no utilizarla.

**Palabras clave:** Generative Artificial Intelligence; risks and potential; Judiciary; human protagonism; governance.

## 1 INTRODUÇÃO

No episódio VI da saga Star Wars (O Retorno do Jedi), o robô C3-PO, um "droide de protocolo" (intérprete de relações-sociais, como ele próprio se qualifica nos filmes), trava uma interessante discussão com um dos protagonistas, Han Solo (Star Wars, 1983):

> C-3PO: I do believe they think I am some kind of god.
> Han Solo: Well, why don't you use your divine influence and get us out of this?
> C-3PO: I beg you pardon, General Solo, but that just wouldn't be proper.
> Han Solo: Proper?
> C-3PO: It's against my programming to impersonate a deity[1].

A escolha desse diálogo para ilustrar o presente artigo, que abordará os dilemas e receios envolvidos na adoção da Inteligência Artificial Generativa - IAG pelo Poder Judiciário não se dá por acaso.

De fato, quando se discute sobre robôs (especialmente humanoides, duas reações automáticas são acionadas na mente humana[2]), em regra: a primeira, oriunda de um desconhecimento sobre as limitações e potenciais da máquina; e a segunda, fruto de uma sensação de perplexidade ante as possibilidades de uma substituição sem precedentes da atuação humana por sistemas computacionais dotados de poderes quase sobrenaturais

[1] Em tradução livre: "C-3PO: Acredito que eles pensam que eu sou algum tipo de deus; Han Solo: Bem, então por que você não usa sua influência divina e nos tira dessa enrascada?; C-3PO: Peço desculpas, General Solo, mas isso simplesmente não seria apropriado; Han Solo: Apropriado?; C-3PO: É contra minha programação personificar uma divindade".

[2] "Humanoides", nesse caso, fazendo referência não à aparência física dos sistemas, mas à experiência do seu usuário/interlocutor.

que trazem um potencial aparentemente ilimitado de intervenção na realidade.

A metáfora trazida pelo diálogo entre C-3PO e Han Solo revela uma dualidade frequentemente observada: a Inteligência Artificial - IA é vista, ora como um instrumento de extrema utilidade, ora como uma entidade.

Em ambos os casos, porém, perde-se de vista um elemento essencial: a responsabilidade e o protagonismo indubitavelmente humanos na programação, na operação e no monitoramento das tecnologias emergentes.

Quando se fala em processos decisórios automatizados, essa dualidade sempre se evidencia, seja nos filmes, seja na realidade.

O Poder Judiciário a vem vivenciando muito intensamente, especialmente após o surgimento da Inteligência Artificial Generativa, havendo aqueles que defendem fortemente sua utilização e muitos que, receando um suposto poder das máquinas, são extremamente refratários a ela, defendendo, inclusive, mediante a edição de normas restritivas, sua proibição ou uso extremamente limitado.

Nessa linha, tendo presentes as reações instintivas de temor e atração que permeiam o discurso sobre IAG, especialmente no âmbito do Poder Judiciário, pretende-se trazer o debate para o campo da racionalidade e dos reais desafios e oportunidades à boa governança institucional, de dados e de processos, na busca da eficiência e do aprimoramento do Sistema de Justiça. Assim, o presente artigo tem dois objetivos:

a) sistematizar conceitos e conhecimentos importantes para a compreensão dos avanços tecnológicos que permeiam o debate sobre o uso da IAG e;
b) investigar de que forma os artefatos e os modelos podem ser empregados, com alguma segurança, no mundo do Judiciário e com aproveitamento de suas potencialidades.

A hipótese que permeia a investigação é a de que os riscos da IAG no sistema de justiça são duplos: eles surgem tanto da aplicação inadequada dessa tecnologia quanto da inércia em não a utilizar.

O foco das reflexões aqui propostas não é um eventual protagonismo descontrolado das máquinas, nem os possíveis erros atribuídos à tecnologia. Tendo presentes os riscos de perda de agência pelos tomadores de decisão, de prejuízos ao componente humano nos julgamentos e à qualidade da produção, o texto investiga de que forma as ameaças se manifestam na prática, avaliando correspondentes alternativas para sua neutralização ou redução.

Explora, também, em que medida os verdadeiros riscos para o Judiciário se situam em ignorar as possibilidades de inovação, transformação e aumento da eficiência que a IAG pode proporcionar e no desconhecimento ou menosvalia que vem sendo atribuída ao protagonismo humano na tomada de responsabilidade com a forma de implementação da ferramenta.

Para isso, com base em pesquisa bibliográfica e documental, será realizado,

inicialmente, um levantamento conceitual sobre Inteligência Artificial e sua evolução histórica, procurando-se abordar, de forma didática, as características da IAG e seus modelos de linguagem.

Posteriormente, serão explorados os principais riscos e receios relacionados ao uso dessa tecnologia no Poder Judiciário, incluindo questões, como alucinações de respostas (respostas equivocadas e/ou desconectadas da realidade), reprodução de vieses preexistentes nos *inputs* utilizados para treinamento, manejo inadequado de dados e informações sensíveis e eventual alienação e perda de autonomia dos juízes. Também será analisado o papel das instituições reguladoras, como o Conselho Nacional de Justiça, e suas iniciativas recentes para regulamentar o uso responsável da IA, propondo-se uma reflexão sobre os custos e riscos da não utilização desse ferramental, diante do volume de dados e de informações, a tornar quase impossível a adequada gestão do conhecimento por meio das tecnologias tradicionais. Ao final, propor-se-á estratégia para equilibrar potencialidades e riscos no contexto do Poder Judiciário.

## 2 NOÇÕES BÁSICAS DE INTELIGÊNCIA ARTIFICIAL, MACHINE LEARNING E IA GENERATIVA

A relação do homem com a natureza e o seu desejo de conseguir manipulá-la para a criação de seres autômatos surgiram bem antes da invenção do primeiro computador. As mais remotas noções de Inteligência Artificial, na verdade, datam da antiguidade, e, segundo Wilkins (2019), a primeira menção efetiva de agentes artificiais com vontade e consciência pode ser rastreada ao mito grego relativo ao gigante autômato de bronze Talos, responsável pela proteção de Creta contra invasores.

A fascinação do homem com o artificial parece ter contaminado o próprio filósofo Renè Descartes, que, encantado com a visita que fez a um jardim autômato em Saint-Germain-en-Laye, Paris, observou que, se um artefato pode ser movido pelo curso das águas, um homem poderia ser movido pela mente, enquanto substância[3] (Wilkins, 2019).

Embora não tão antigos quanto esses frutos da imaginação humana, os fundamentos matemáticos para a Ciência da Computação e, consequentemente, para a Inteligência Artificial já existem há mais de dois séculos.

A álgebra booleana, consistente no mesmo conceito de lógica binária utilizado por computadores hoje em dia (0 ou 1, verdadeiro ou falso), foi desenvolvida por George Boole ainda em 1854. Antes mesmo de Boole, por volta de 1822, o matemático inglês Charles Babbage fez o desenho de um dos primeiros computadores, o "Difference Engine", uma máquina analítica de propósito geral, que, se desenvolvida, teria o

[3] Outras referências sobre Descartes e seres autômatos em geral podem ser encontradas no excelente artigo da professora de história da Universidade de Stanford Jessica Riskin, no endereço eletrônico: https://arcade.stanford.edu/rofl/machines-garden

equivalente a 1 kilobyte de memória; pouco para hoje em dia, mas uma enormidade na época (Wilkins, 2019).

Os avanços subsequentes dos séculos XIX e XX, desde os computadores voltados a tarefas específicas, passando pela Segunda Grande Guerra, até aqueles generalistas (universais) que vieram depois, pavimentaram o caminho para a Inteligência Artificial encontrada hoje em dia e sonhada por Alan Turing, nas décadas de 1930 e 1940.

É, inclusive, creditado a Alan Turing, um matemático, lógico e cientista de computação britânico, o primeiro experimento que daria origem à ciência moderna da computação. Sua principal contribuição nesse campo foi o artigo "On Computable Numbers, with an application to the Entscheidungsproblem" (Turing, 1936), em que descrevia uma máquina teórica que poderia resolver qualquer problema tão logo este fosse codificado nas fitas de instrução (*paper tape instructions*) apropriadas.

Na década de 1930, Turing (1936 p. 13) afirmou ser possível inventar uma máquina "which can be used to compute any computable sequence. If this machine U is supplied with a tape on the beginning of which is written the S.D of some computing machine M. Then U will compute the same sequence as it"[4]. Em outras palavras, um computador poderia rodar dentro de outro computador. Essa propriedade ficou posteriormente conhecida *como Turing-completeness* (Michaelson, 2020).

As contribuições de Alan Turing para a ciência computacional não param por aí. Suas percepções fundamentais para o desenvolvimento da Inteligência Artificial podem ser encontradas em outro de seus artigos "Computing Machinery and Intelligence", de 1950; no qual ele prevê um computador poderoso o suficiente para simular a inteligência humana. Nesse artigo, ele defende a existência eventual de um computador inteligente e afasta diversos argumentos que são aventados quanto a tal possibilidade (desde argumentos de ordem teológica, quanto de ordem mais técnica e matemática) (Turing, 1950).

Também no referido texto, Turing (1950) formula, logo em sua introdução, o famoso "Turing Test" (Teste de Turing), como forma de quantificar o quão inteligente é o computador.

Logo após Turing, no final da década de 1950, e durante toda a década de 1960, pesquisadores, como Marvin Minsky, Allen Newell, Herbert Simon e John McCarthy, passaram a desenvolver as primeiras experiências com redes neurais artificiais (Wilkins, 2019), e a inteligência computacional pareceu algo cada vez mais concreto no universo dos computadores e dos algoritmos codificados.

Atualmente, sistemas experts, especialmente generativos, assombram por sua

[4] Em tradução livre: "é possível inventar uma única máquina que pode ser usada para computar qualquer sequência computável. Se esta máquina U for alimentada com uma fita no começo da qual está escrito um número satisfatório de alguma máquina computacional M, então U computará a mesma sequencia que M".

suposta capacidade de superar a inteligência e a criatividade humanas, e perguntas sobre as potencialidades da máquina inteligente permeiam as discussões também entre os leigos.

Porém, o que exatamente é a Inteligência Artificial que se discute hoje em dia e é o centro deste artigo?

Quando o foco é Inteligência Artificial- IA, normalmente as pessoas se referem a dois tipos: *Artificial General Intelligence - AGI* (Inteligência Artificial Geral) e *Narrow AI* (Inteligência Artificial Fraca ou Limitada).

A AGI é, para usar uma expressão do autor Wilkins (2019), o "santo graal" da pesquisa computacional. Seria uma máquina com inteligência comparável à humana, no sentido de que conseguiria cumprir tarefas a partir da generalização de técnicas de resolução de problemas, projetadas para questões específicas.

Um computador equipado com Inteligência Artificial Geral compreenderia a linguagem em um nível muito fundamental, e suas habilidades ingressariam em campos da elaboração inteligente e do conhecimento humanos de forma não imaginada por seu criador (um bom exemplo desse tipo de tecnologia seria o HAL 9000, do filme "2001: uma odisseia no espaço"[5] ou, mais recentemente, o Winston[6], do livro "A Origem", de Dan Brown).

Assim, quando se fala em Inteligência Artificial atualmente em aplicação, seja em processos decisórios automatizados, carros autônomos, modelos largos de linguagem – LLMs (como ChatGPT e Claude), "assistentes virtuais" (como Alexa e a Siri) ou computadores que jogam xadrez (como Deep Blue ou AlphaGo), a referência é o segundo tipo específico de IA, a *Narrow AI* (Inteligência Artificial Fraca ou Limitada).

A IA Limitada se divide em diversas categorias, mas o foco inicial aqui será a mais conhecida delas, a chamada de *Machine Learning,* ou aprendizado da máquina, que é a técnica que está atualmente sendo aplicada dentro do Sistema de Justiça não só no Brasil, mas também, nos Estados Unidos (O'Neil, 2016) e na Estônia (E-Estonia, 2020), para citar alguns exemplos.

*Machine learning* é um processo no qual algoritmos podem "aprender" a partir de

[5] Tentando não dar *spoiler* de uma excelente obra cinematográfica, HAL é um computador instalado a bordo da nave espacial Discovery e responsável por todo o seu funcionamento. Ele, representado no filme por diversas câmeras espalhadas pela nave, é capaz de falar espontaneamente, realizar reconhecimento facial, fazer leitura labial, apreciar manifestações artísticas, interpretar e expressar emoções, raciocinar e jogar xadrez. Ao longo da saga narrada na trilogia (2001: uma odisseia no espaço, 2010: uma odisseia no espaço 2 e 3001: a odisseia final), HAL demonstra pensar além da sua programação, apresenta arrependimento, medo e também propósito de se sacrificar em prol de outros, a despeito de receber ordem de não o fazer.

[6] Winston é um sistema de Inteligência Artificial também, como HAL, não humanoide, que demonstra conhecimento da linguagem em seu nível mais básico e capacidade de fazer associações que vão bem além de um sistema expert. Winston "aprende" espontaneamente a partir de buscas autônomas na Internet, utiliza improvisações constantes, encontra soluções para além da sua programação e é um dos grandes protagonistas da obra, chegando a ser confundido com um humano diversas vezes.

uma grande quantidade de dados colocados dentro de um sistema, em vez de serem *softwares* que simplesmente percorrem um caminho ou caminhos predeterminados pelo seu "criador".

Os computadores baseados em *Machine Learning* são treinados para executar uma tarefa e aprender com essa execução, ultrapassando, ao final, as primeiras instruções dadas por seu programador.

Essa técnica de IA é bastante poderosa, como se pode verificar; porém tem limites e, por isso, pertence à gama de Inteligência Artificial Limitada. Por meio do uso de *Machine Learning*, os sistemas podem aprender a fazer algo (como elaborar projetos de decisões judiciais baseadas em precedentes) muito bem; porém não conseguem generalizar esse aprendizado para o alcance de soluções outras, como um homem o faria (por exemplo, passar a jogar xadrez a partir de informações retiradas da Internet), a não ser que sejam expressamente "(re)orientados" e "(re)treinados" para isso. Eles possuem, literalmente, uma inteligência restrita e são altamente dependentes de *inputs* humanos deliberados.

Dados são alimentados no sistema, um algoritmo é selecionado, hiperparâmetros são configurados e ajustados, e a máquina se prepara para conduzir análises das variáveis insertas. O computador começa a decifrar relações encontradas nos dados por meio de um processo de tentativa e erro, tudo isso visando à predição de valores futuros (Theobald, 2017).

Embora haja uma relação entre programador e máquina, os dois operam em um nível diferente da programação tradicional, em que os resultados são predefinidos. Isso acontece porque, após a inserção do código pelo homem, a máquina passa a, literalmente, "aprimorar" os algoritmos alimentados com novas perspectivas, a partir de conexões deliberadas entre os dados fornecidos.

Um exemplo interessante desse "aprimoramento" é o caso do Alpha Go, programa de computador desenvolvido pela Alphabet Inc's Google Deep Mind e descontinuado em 2017, que usava a tecnologia de redes neurais. Alimentado apenas com as regras do xadrez, Alpha Go literalmente aprendeu o jogo a partir da prática, jogando com humanos e outros computadores (Hölldobler; Möhle; Tigunova, 2025).

Comumente, *Machine Learning* é confundida com automação. Porém, enquanto, na primeira, há uma verdadeira (embora limitada) Inteligência Artificial, na análise de dados e na tomada de decisão; na segunda, há apenas um algoritmo simples que determina que, em um certo ponto, uma tarefa específica será realizada pelo sistema e não pelo homem. Trata-se, nesse caso, apenas de uma programação inteligente e eficaz.

No caso do aprendizado da máquina, o sistema formula, a partir dos dados insertos, decisões autônomas, emulando o processo decisório humano e elevando a relação entre programador e máquina a outro nível de complexidade, se comparado com

a programação tradicional (Theobald, 2017).

Nesse passo, as tarefas envolvidas em *Machine Learning* podem ser subdivididas em duas classes diferentes: aprendizado supervisionado (*supervised learning*) e aprendizado não supervisionado (*unsupervised learning*).

No aprendizado supervisionado, o programador ou pesquisador utiliza dados que já foram categorizados e classificados para auxiliar o processo de aprendizado.

O aprendizado não supervisionado, por outro lado, não usa dados pré-classificados. Algoritmos de aprendizado não supervisionado são utilizados para explorar uma estrutura de dados (por meio de regressão, classificação, agrupamento, filtragem colaborativa ou aprendizagem forçada) e encontrar relações entre eles.

A vantagem do aprendizado não supervisionado é que ele permite que se descubram padrões nos dados que não foram previamente identificados; e é dessa identificação que surgem soluções e decisões não visualizadas anteriormente pelo homem.

Por essas inúmeras possibilidades de tratamento e associação de dados e informações, *Machine Learning*, como já se mencionou, é a forma de Inteligência Artificial que está sendo utilizada com mais frequência no mundo do Direito.

Nesse contexto, a IA generativa é uma aplicação ainda mais sofisticada do aprendizado de máquina, conhecida como aprendizado profundo (*Deep Learning*), que se apoia em redes neurais artificiais complexas para identificar padrões em grandes volumes de dados e, com base em comandos do usuário gerar, ao final, conteúdos inéditos (Banh; Strobel, 2023).

O termo "generativa" se refere justamente à sua capacidade de gerar esses resultados novos e coerentes, simulando o comportamento criativo humano em determinadas tarefas. Exemplos conhecidos dessa tecnologia incluem modelos de linguagem de grande escala (LLMs), utilizados em ferramentas, como ChatGPT, Claude, Gemini, Deep Seek e geradores de imagens, como o DALL-E. Tais sistemas não apenas reconhecem e interpretam padrões como as formas mais básicas de *Machine Learning*, mas também criam produtos inéditos (textos, imagens, sons, códigos de programação etc) com base nos padrões aprendidos e inputs recebidos durante o seu treinamento/aprendizado.

Apesar de sua sofisticação, a IAG, como forma de IA limitada, não possui compreensão profunda dos conteúdos que produz nem capacidade de generalização para além do domínio específico em que foi treinada.

Na verdade, ela opera, fundamentalmente, por meio da constatação da alta probabilidade de determinadas sequências textuais ou contextuais serem formadas dentro de um determinado assunto ou domínio do conhecimento e da replicação dessas sequências no produto que oferta (Mollick, 2024); ou seja, mediante meras correlações estatísticas entre os padrões textuais ou contextuais identificados nos seus dados de treinamento (Mollick, 2024).

Assim, os sistemas já citados, embora aparentemente pareçam coerentes e prolixos, nada mais fazem do que associar de forma dinâmica palavras que estaticamente costumam aparecer juntas. Eles não possuem compromisso com a coerência ou a consistência, apenas com as correlações probabilísticas que encontram.

Em resumo, a IAG pode, por exemplo, redigir uma peça, criar uma narrativa ou responder a perguntas de forma detalhada, mas não tem "intenção" ou "consciência". Todo seu desempenho depende de padrões extraídos e probabilidades calculadas com base nos *inputs* que recebeu durante o seu aprendizado, aliados aos *prompts*/comandos do usuário, e nada além disso.

## 3 REFLEXÕES ATUAIS SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO

A crescente incorporação de tecnologias de Inteligência Artificial (IA) no Poder Judiciário tem se consolidado como um fator determinante na modernização da prestação jurisdicional (Santos, 2025), prometendo ganhos notáveis em eficiência, celeridade e otimização da gestão processual (Rodrigues *et al.*, 2024). Contudo, a par das grandes expectativas, a implementação de tais ferramentas, notadamente as de IA generativa, suscita desafios éticos e jurídicos complexos, que demandam uma reflexão aprofundada sobre a compatibilidade entre a inovação tecnológica e os princípios fundamentais que regem o Estado Democrático de Direito (Gabriel; Porto; Araujo, 2025).

Um dos riscos mais debatidos é a chamada "mecanização excessiva das decisões" (Santos, 2025), que consiste na possibilidade real de a automação conduzir a julgamentos padronizados, sem a devida análise contextualizada e subjetiva que a aplicação do Direito exige (Santos, 2025).

Como defendem Gabriel, Porto e Araujo (2025), a atividade jurisdicional envolve elementos que transcendem a mera operação lógica, demandando interpretação, sensibilidade e ponderação de valores que, ao menos por enquanto, não podem ser adequadamente capturados por algoritmos. A essência do Direito, permeada por valorações extraídas da experiência humana, refletiria, para esses autores, a diversidade cultural e histórica que compõe o tecido social, revelando-se tanto uma arte quanto uma ciência (Gabriel; Porto; Araujo, 2025). Sobre o assunto, Bezerra Neto (2025) sustenta que "a capacidade de valorar subjetivamente, conectando emoções e racionalidade, é o que distingue fundamentalmente o juízo humano da análise probabilística realizada por inteligências artificiais".

Nesse sentido, a decisão final, com toda a sua carga valorativa, deveria permanecer como uma prerrogativa exclusivamente humana, consagrando o que se tem chamado de "reserva de humanidade" (Gabriel; Porto; Araujo, 2025).

Intimamente ligada a essa questão está a falta de transparência de muitos sistemas

de IA, que operam como verdadeiras "caixas-pretas", cujo funcionamento é de difícil compreensão, inclusive para seus desenvolvedores (Santos, 2025).

Essa opacidade algorítmica iria de encontro a um dos princípios basilares do Direito: a necessidade de fundamentação clara e compreensível das decisões judiciais (Bonat; Vale; Pereira, 2023).

Para que o uso da IA fosse legítimo, seria imperativo que se garantisse a explicabilidade dos seus resultados, permitindo que magistrados, as partes e a sociedade compreendessem os critérios utilizados pelo sistema (Bonat; Vale; Pereira, 2023). A ausência de transparência não apenas minaria a confiança pública, mas também inviabilizaria o controle de legitimidade das decisões e o exercício pleno do contraditório. A preocupação com a falta de transparência é tão presente que, em pesquisa realizada pelo Conselho Nacional de Justiça, a maioria dos servidores e assessores indicou não revelar o uso de IAG a seus pares ou superiores hierárquicos, o que dificulta os processos internos de revisão e pode levar a imprecisões na prestação do serviço jurisdicional (CNJ, 2024).

Outro desafio crucial residiria na presença de vieses algorítmicos que têm o potencial de comprometer a imparcialidade das decisões e que serão discutidos com mais vagar nos tópicos seguintes deste trabalho. Um sistema de IA treinado com base em decisões pretéritas pode, em tese, acabar por reproduzir e até amplificar padrões discriminatórios relacionados a gênero, raça ou origem social, resultando em decisões injustas (Santos, 2025).

Diante de tais complexidades, a construção de uma moldura robusta de governança, pautada na ética e na supervisão humana, torna-se um pilar fundamental para a implementação responsável da IA. A regulamentação do tema no Brasil, como a Resolução n. 615/2025 do Conselho Nacional de Justiça, que sucedeu a Resolução 332/2020 e será brevemente tratada mais adiante, buscou, sobre o assunto, estabelecer um equilíbrio, definindo modelos de interação que evitem uma dependência irrestrita e acrítica da tecnologia, o chamado "Parasitismo Lex Machinae" (Divino, 2025), que poderia levar a julgamentos injustos e marginalizar o ser humano.

## 4 IA GENERATIVA E RISCOS DE SUA NÃO COMPREENSÃO ADEQUADA: ALUCINAÇÕES DE RESPOSTAS, REPRODUÇÃO DE VIESES PREEXISTENTES E MANEJO INADEQUADO DE DADOS E INFORMAÇÕES SENSÍVEIS

A rápida popularização da Inteligência Artificial Generativa no Poder Judiciário como a tecnologia emergente do momento tem suscitado discussões prementes não apenas em relação às questões jurídicas e éticas, como exposto no tópico anterior, mas também, acerca dos riscos e dos desafios inerentes à sua utilização, especialmente pela falta de uma devida compreensão de suas limitações técnicas e implicações práticas pelo

público em geral.

Como mencionado, como IA limitada, a IAG precisa ser compreendida como restrita ao que se propõe – gerar textos, imagens, sons etc. – e não ser vista como uma plataforma pensante com intencionalidade para além da criação de conteúdo a partir de parâmetros (ainda que imensuráveis) a que tem acesso por meio de diversas fontes de dados.

## 4.1 ALUCINAÇÕES E VIÉS DE AUTOMAÇÃO

Um dos aspectos mais críticos do uso não informado da IAG, que tem o potencial de comprometer a legitimidade e a segurança dos seus resultados, refere-se ao fenômeno das "alucinações" de respostas (Hatem; Simmons; Thorbton, 2023)[7] – termo criado para designar a geração de conteúdo incorreto ou inexistente, que apresenta, porém, para o usuário desavisado, aparência de veracidade.

Esse comportamento dos sistemas generativos, com frequência, decorre da arquitetura dos próprios modelos de linguagem de grande escala (Large Language Models - LLMs), que, como já se apontou, não possuem real compreensão semântica ou capacidade de raciocínio, embora emulem com extraordinária similitude essas habilidades tão humanas.

A questão torna-se particularmente sensível em áreas do conhecimento mais subjetivas e relacionadas a um raciocínio lógico e indutivo mais apurado, como no Direito. A aparente coerência e fluência do texto gerado pode induzir o usuário desprevenido ao que se denomina "viés de automação" (*automation bias*) ou ao viés da complacência (*automation complacency*). O primeiro, uma tendência à confiança excessiva (e, por vezes, "cega") no resultado apresentado pela máquina (Bahner; Hüper; Manzey, 2008), que se presume, não teria razões para mentir. O segundo, uma tendência do usuário, em um determinado momento, a tornar-se tão confortável e acomodado com sistemas automatizados que passa a deixar de lado sistemas de checagem e conferência anteriormente usados (Parasuraman; Manzey, 2010).

Segundo Parasuraman e Manzey (2010), a complacência e o viés de automação representam diferentes manifestações de fenômenos sobrepostos induzidos pela automação, com a atenção desempenhando um papel central, resultado de uma interação dinâmica de características pessoais e situacionais.

Tome-se como exemplo um cenário em que magistrados utilizam sistemas de inteligência artificial para auxílio na análise de pedidos de liberdade (seja em audiências

[7] Optou-se pela utilização já bastante difundida do termo "alucinações" para descrever esse específico fenômeno no funcionamento da IAG. No entanto, é importante consignar que a palavra tem sido combatida por alguns profissionais que trabalham no cuidado da saúde mental, conforme esclarecem, por diversos outros.

de custódia, seja para avaliação da possibilidade de adoção de medidas cautelares). O sistema, alimentado por uma vasta base de dados de decisões judiciais pretéritas, apresenta recomendações aparentemente bem fundamentadas sobre a concessão ou não do benefício.

Diante de um caso concreto envolvendo um furto simples, o magistrado, mesmo diante de elementos singulares que poderiam justificar a concessão da liberdade - como primariedade, vínculos empregatícios estáveis e residência fixa -, mas já assoberbado de casos semelhantes, poderá vir a acolher, especialmente diante de um comando (*prompt*) de menor complexidade, a envolver poucas variáveis, a sugestão automatizada de manutenção da prisão preventiva, baseada meramente em correlações estatísticas extraídas do histórico decisório.

Por fim, além de o ser humano ter o potencial de amplificar as alucinações por meio de enviesamentos internos (sejam relacionados à confiança cega na automação sejam aqueles advindos da complacência e do conforto cognitivo relacionado), é também o usuário final - e não o desenvolvedor ou a ferramenta – o maior responsável por alguns dos principais gatilhos que ensejam as alucinações e que são oriundos primordialmente de uma falta de conhecimento sobre a IAG e suas plataformas.

Nessa linha, pode-se citar eventual ambiguidade dos comandos, que frequentemente resulta em interpretações imprecisas pelo sistema e o força a fazer suposições ou preencher lacunas com informações que podem não ser corretas para o contexto específico. Por exemplo, solicitações genéricas – e bem comuns no mundo jurídico - como "analise o documento" ou “resuma o caso”, não fornecem parâmetros suficientes para que o sistema compreenda exatamente qual tipo de análise deve ser realizada ou quais informações são relevantes no resumo que foi pedido.

Da mesma forma, a não especificação de forma adequada do escopo temporal ou contextual da consulta, com perguntas sem delimitação clara de período ou área do conhecimento, o uso inadequado de terminologia técnica ou a mistura de conceitos de diversas áreas sem a devida contextualização pode confundir o sistema ou levá-lo a combinar informações de diferentes contextos, resultando em respostas que, embora pareçam coerentes, não o são.

Por exemplo, caso se pergunte à IAG "como funciona o processo de adoção?", sem especificação de país ou tempo, ela pode combinar regimes legais de diversos países ou misturar procedimentos antigos com práticas atuais, criando um conteúdo que, embora pareça completo, pode conter inconsistências ou anacronismos.

Ou se a pergunta for: "explique como funciona o recurso no processo e quais são seus efeitos", o resultado poderia ser um texto que, embora pareça bem estruturado e abrangente, na verdade, teria o enorme potencial de apresentar inconsistências técnicas graves e misturas inadequadas de diferentes ramos do direito processual, potencialmente levando a interpretações errôneas sobre como os recursos realmente funcionam em cada

contexto específico.

Além disso, a ausência de protagonismo humano no processo de tomada de decisão é um importante risco a ser considerado. O manejo irresponsável e não supervisionado da IAG, com a delegação do poder decisório, pode ser, ao mesmo tempo, causa e consequência de uma alienação e perda de agência (*loss of agency*) dos usuários, o que, em se tratando de juízes, pode levar a graves prejuízos aos direitos dos jurisdicionados (Münch; Ferraz, 2024).

Esse risco é especialmente presente quando se pretende dar tratamento ao excesso de litigiosidade e cumprir metas de produtividade, mediante o uso massivo de ferramentas tecnológicas, e resulta agravado quando se percebe que grande parte das ferramentas de IA já desenvolvidas no Judiciário centra foco exclusivamente no aumento da eficiência na solução dos casos judicializados (Münch; Ferraz, 2024).

O panorama se torna ainda mais complexo quando consideramos a interação entre esses diferentes fatores de risco. Por exemplo, a combinação entre alucinações e vieses pode resultar na geração de conteúdo não apenas factualmente incorreto, mas também discriminatório, enquanto a alienação, a perda de agência e o manejo inadequado de dados sensíveis podem ser agravados pela tendência à confiança excessiva no sistema.

A falta de compreensão sobre as limitações e as capacidades reais da IA também pode contribuir para as alucinações. Muitos usuários assumem erroneamente que o sistema possui conhecimento universal e atualizado sobre todos os assuntos (como uma poderosa ferramenta de pesquisa), quando, na realidade, como já se mencionou, seu conhecimento é limitado aos dados de treinamento e ao limite temporal de acesso à base de dados dinâmica. Isso pode resultar em consultas que excedem as capacidades da IAG ou que exigem informações além de seu escopo de conhecimento.

Por fim, a ausência de *feedback* adequado ao sistema sobre possíveis erros ou inconsistências impede o aprimoramento contínuo das interações. Quando os usuários não reportam ou identificam adequadamente as alucinações, perde-se a oportunidade de desenvolver mecanismos mais robustos de prevenção e correção desses problemas.

## 4.2 AMPLIAÇÃO DOS VIESES

Outro aspecto crítico no uso de IAG diz respeito à reprodução e à potencial ampliação de vieses pré-existentes nos dados utilizados para treinamento/aprendizado da máquina (por outros, Buolamwini, 2017 e Santos, 2025). Esse fenômeno pode manifestar-se de formas sutis ou explícitas, perpetuando discriminações relacionadas a gênero, raça, origem social e étnica, orientação sexual ou outras características presentes nos *inputs* recebidos.

A complexidade da questão é aumentada pela natureza dos modelos generativos, caracterizados pela incerteza e pela imprevisibilidade na construção dos seus resultados,

que torna difícil a identificação tempestiva e eficiente destes vieses. Ademais, a própria estrutura do processo de treinamento, baseada em maximização de probabilidades de combinações textuais, pode levar ao aumento de padrões discriminatórios presentes nos dados originais, em um círculo vicioso de confirmação e reconfirmação.

Para ilustrar tal problemática, considere-se o caso dos sistemas de reconhecimento facial implementados em contextos de segurança pública. Tais sistemas, quando treinados predominantemente com bases de dados compostas por faces de determinados grupos étnico-raciais - historicamente, indivíduos caucasianos -, apresentam taxas significativamente mais elevadas de falsos positivos ao analisar faces de pessoas negras ou asiáticas, diante da incapacidade de discernir traços fenotípicos com a mesma acurácia com que interpreta os dados com que foi treinada. Em um cenário de utilização desses sistemas para identificação de suspeitos em investigações criminais, esta disparidade técnica transmuta-se em grave questão de justiça social, podendo resultar em abordagens policiais equivocadas e, até mesmo, em prisões injustas, reforçando e amplificando padrões discriminatórios já existentes no sistema de justiça criminal[8].

## 4.3 MANIPULAÇÃO DE DADOS SENSÍVEIS

A gestão inadequada de dados sensíveis representa outro fator relevante de risco. A facilidade de uso e a interface interativa das ferramentas de IAG podem induzir à inclusão inadvertida de dados confidenciais ou pessoais nas entradas. Esta prática pode resultar em violações da Lei Geral de Proteção de Dados - LGPD e de outras legislações correlatas, potencialmente expondo dados sensíveis a processamentos não autorizados, com inúmeras possíveis consequências.

Em nível internacional, merece destaque, recentemente, o caso da *startup* chinesa de inteligência artificial DeepSeek, que está no centro de uma controvérsia após o vazamento de dados que expôs informações sensíveis de usuários. O caso levanta preocupações sobre violações de privacidade e possíveis infrações de direitos de propriedade intelectual, especialmente diante das acusações feitas pela OpenAI.

A empresa de cibersegurança Wiz revelou que a DeepSeek deixou exposto um banco de dados contendo mais de um milhão de registros, incluindo históricos de conversas dos usuários, chaves de API e detalhes internos do sistema. Esses dados estavam desprotegidos e acessíveis publicamente, o que representa uma grave falha de segurança (Nagli, 2025).

[8] Registre-se, a propósito, que o Comitê para a Eliminação da Discriminação Racial da Organização das Nações Unidas – ONU (2020), em sua Recomendação Geral n. 36, de 24 de novembro de 2020 observou que o aumento do uso pelas forças policiais de algoritmos, inteligência artificial, reconhecimento facial e de outras tecnologias aumenta os riscos de aprofundamento do racismo, da discriminação racial e da xenofobia.

Além disso, de acordo com uma análise conduzida pela *Reuters* (Yim, 2025), a DeepSeek coleta extensivamente dados pessoais dos usuários e os armazena em servidores na China. Segundo a política de privacidade da empresa, informações como entradas de texto ou áudio, arquivos enviados e históricos de chat são coletados e podem ser compartilhados com autoridades chinesas, em conformidade com as leis locais. O Serviço Nacional de Inteligência da Coreia do Sul classificou essa prática como "excessiva e preocupante", destacando os riscos para a segurança digital dos usuários.

Governos de diversos países começaram a reagir ao problema. Itália, Coreia do Sul e Taiwan já adotaram medidas para restringir ou investigar o uso do DeepSeek em seus territórios (DeepSeek, 2025). Especialistas em segurança alertam que esse incidente reforça a necessidade de regulamentações mais rígidas sobre como empresas de inteligência artificial tratam os dados de seus usuários (Límon, 2025).

Além das preocupações com privacidade, a OpenAI acusou a DeepSeek de violar direitos autorais ao utilizar seus modelos proprietários sem permissão para treinar um chatbot concorrente. Segundo um relatório publicado pelo *The Conversation* (2025), a OpenAI sustenta que a DeepSeek empregou técnicas de "destilação", nas quais um modelo menor aprende a imitar o comportamento de um modelo maior e mais complexo. Esse processo pode permitir que empresas concorrentes desenvolvam produtos semelhantes sem investir no desenvolvimento completo de um modelo original.

Essa questão levanta um debate mais amplo sobre os direitos de propriedade intelectual na área da inteligência artificial. O World Economic Forum (2024) destacou que, com a crescente adoção de modelos de IA de código aberto, muitas empresas estão explorando brechas legais para replicar tecnologias desenvolvidas por grandes *players* do setor. No entanto, a OpenAI também enfrenta questionamentos sobre a legalidade do uso de dados protegidos por direitos autorais para treinar seus próprios modelos de IA, o que a coloca em uma posição delicada ao acusar outras empresas de condutas semelhantes (The Conversation, 2025).

Diante do ocorrido, o caso da DeepSeek reforça a necessidade de regulamentações mais robustas para proteger a privacidade dos usuários e garantir que os direitos de propriedade intelectual sejam respeitados no desenvolvimento de tecnologias de inteligência artificial. Empresas e governos precisam trabalhar juntos para estabelecer regras claras que promovam inovação sem comprometer a segurança digital e a ética no uso de dados.

Nesse sentido, o desenvolvimento de rigorosos processos de validação, monitoramento contínuo e auditorias regulares humanas constitui pilar fundamental para garantir que os parâmetros éticos e legais estabelecidos sejam efetivamente observados na prática. Estes processos devem ir além de verificações técnicas superficiais, incorporando metodologias sofisticadas de avaliação de impactos nos direitos fundamentais, análise de vieses discriminatórios e verificação da conformidade com

princípios constitucionais.

O monitoramento contínuo deve ser estruturado como um sistema de vigilância ativa que detecte precocemente desvios, anomalias ou impactos não antecipados. Isso inclui o desenvolvimento de indicadores de performance ética, métricas de coerência algorítmica e mecanismos de *feedback* dos usuários do sistema.

Por outro lado, a capacitação contínua e adaptativa para garantir o protagonismo humano, acompanhando a evolução tecnológica e incorporando lições aprendidas da experiência prática, representa elemento crucial para evitar tanto os vieses quanto as causas humanas de alucinações nas respostas da IAG. Essa capacitação deve abordar não apenas aspectos técnicos sobre o funcionamento dos sistemas de IA, mas também desenvolver competências importantes para avaliação da qualidade, adequação e limitações dos *outputs* gerados.

## 5 PODER JUDICIÁRIO BRASILEIRO E IA GENERATIVA

Segundo o Relatório de Pesquisa "O uso da Inteligência Artificial Generativa no Poder Judiciário Brasileiro" (CNJ, 2024), produzido pelo Conselho Nacional de Justiça, o uso de ferramentas de Inteligência Artificial Generativa (IAG) tem se expandido nas atividades judiciais e administrativas dos tribunais brasileiros, sendo geralmente voltado para a geração de textos, aperfeiçoamento textual, tradução e resumo de documentos, além de criação de tabelas e sistematização de informações.

O ChatGPT aparece como a ferramenta mais utilizada, seguido por outras como Copilot, Gemini e Bing AI.

Embora o uso dessas ferramentas esteja claramente em uma curva crescente, o relatório apontou que ele vem acompanhado de preocupações também crescentes. Nessa linha, as principais dificuldades relatadas pelos usuários estão na falta de familiaridade com as ferramentas (51,9%), na falta de confiança nos resultados gerados (48,9%) e nas preocupações quanto à licitude do uso (cerca de 15%).

Ainda mais, ainda que metade dos magistrados (49,4%) e servidores (49,5%) do Poder Judiciário já tenha tido experiência com ferramentas de Inteligência Artificial Generativa (IAG), elas foram utilizadas, em regra, em plataformas abertas disponíveis na internet e sem o conhecimento ou chancela institucional, o que pode dificultar processos internos de revisão e supervisão do conteúdo gerado e controle dos dados e informações sensíveis compartilhados.

Em resposta a esses achados, como já mencionado, o Conselho Nacional de Justiça publicou recentemente, em março de 2025, a Resolução nº 615/2025 (Brasil, 2025). Esse normativo, que revoga a anterior Resolução nº 332/2020, estabelece diretrizes específicas e abrangentes para o desenvolvimento, utilização e governança de soluções de IA, com especial atenção à inteligência artificial generativa.

Um dos aspectos mais relevantes da Resolução 615/2025 é a adoção de uma abordagem baseada em riscos, classificando as aplicações de IA em categorias de alto risco, baixo risco ou risco excessivo (vedado). As soluções de alto risco - como aquelas destinadas à "aferição da adequação dos meios de prova e sua valoração" ou "formulação de juízos conclusivos sobre a aplicação da norma jurídica" - estão sujeitas a requisitos mais rigorosos, incluindo avaliação de impacto algorítmico e supervisão contínua.

Por outro lado, aplicações de baixo risco, como "execução de atos processuais ordinatórios" ou "produção de textos de apoio para facilitar a confecção de atos judiciais", possuem requisitos de governança mais simplificados, embora ainda exijam supervisão humana adequada.

Por fim, dentre outros aspectos, a resolução dedica capítulo específico ao uso de modelos de linguagem de larga escala (LLMs) e sistemas de IAG, estabelecendo diretrizes tanto para uso corporativo quanto individual. Magistrados e servidores, assim, podem utilizar ferramentas comerciais de IAG, desde que observadas condições como capacitação adequada, uso auxiliar (vedada a tomada autônoma de decisões), proteção de dados sensíveis e conformidade com a LGPD.

Quando os tribunais contratam soluções de IAG, a resolução exige que as empresas fornecedoras se comprometam com o respeito à legislação brasileira, incluindo a vedação do uso de dados do Poder Judiciário para treinamento não autorizado e a implementação de mecanismos de *privacy by design* (privacidade desde a concepção) *e privacy by default* (privacidade por padrão)[9].

Entre os desafios à frente, no uso da IAG, no Poder Judiciário, está a necessária percepção de que seus artefatos e suas funcionalidades poderão contribuir para que se alcancem resultados muito além do aumento facilitado da produtividade e da celeridade nas unidades judiciárias. Trata-se de um caminho que se abre à expansão da capacidade de tomar decisões, diante de suas infinitas possibilidades de cruzamento de informações e de geração de *insights*.

Atualmente, porém, parece que é apenas a busca pela eficiência, traduzida em fazer mais com menos recursos, que está fomentando a utilização da IAG. O movimento eficientista do Judiciário, fortemente influenciado pelas metas de produtividade, parece ter como único ou principal propósito, domar o absurdo volume de processos, quando o propósito da existência de um sistema de justiça é muito maior e mais profundo.

É interessante perceber a espiral de sucesso que tende a surgir, quando se

[9] *Privacy by design* significa que a proteção de dados deve ser incorporada desde o início do desenvolvimento de qualquer sistema ou serviço, não como uma adaptação posterior, enquanto *Privacy by default* estabelece que as configurações mais restritivas de privacidade devem ser aplicadas automaticamente, coletando apenas dados essenciais pelo menor tempo necessário (Brasil, 2024, p. 11-13). Em ambos os casos, a privacidade deixa de ser opcional para se tornar um direito fundamental integrado ao design tecnológico desde a concepção.

implementam inovações, que também gera uma espécie de efeito alucinatório entre os usuários. No caso, a euforia que decorre do aumento quase que instantâneo e antes impensável nos índices de produtividade, tende a encobrir os possíveis paraefeitos da solução implantada, sobre a própria litigiosidade, revelada nos índices de judicialização e recorribilidade.

O alívio dos sintomas de um problema complexo faz pensar que ele foi resolvido, já que o comportamento geral melhora antes de piorar. O *feedback*, a informação que volta para o sistema, tende a chegar com defasagem (Senge, 2013), o que dificulta, no curto prazo, perceber os sinais de que as causas do problema não foram alcançadas, quiçá estejam sendo retroalimentadas. Esse efeito alucinatório faz com que, diante do aumento de processos e recursos, se continue a investir em aumento de produção, já que, em razão da defasagem do feedback, não se associa o problema de hoje às soluções ontem adotadas e a verdade é que a litigiosidade tem empurrado de volta (Ferraz, 2023). A esse cenário, soma-se o chamado paradoxo da eficiência ou efeito bumerangue (Jevons, 1865), que descortina o fato de que o uso eficiente de um produto ou serviço contribui para aumentar o referido consumo.

## 6 BENEFÍCIOS QUE PODEM SURGIR PELO USO DA IA GENERATIVA NO PODER JUDICIÁRIO

Apesar dos riscos e dos desafios inerentes ao uso da Inteligência Artificial generativa no contexto do Poder Judiciário, é essencial reconhecer que sua implementação responsável pode trazer benefícios significativos para o Sistema de Justiça brasileiro. Quando adequadamente compreendida, governada e supervisionada por operadores qualificados, a IAG pode tornar-se uma ferramenta valiosa para o aprimoramento da prestação jurisdicional.

Um dos seus benefícios mais imediatos reside na redução do tempo necessário para a análise de informações complexas ou muito volumosas. Em um cenário no qual magistrados e servidores frequentemente lidam com processos que envolvem milhares de informações e documentos, a capacidade da IA de sintetizar, organizar e destacar informações relevantes pode representar um ganho substancial de eficiência.

Essa otimização na gestão do conhecimento que transita ou é gerado no Poder Judiciário não deve ser compreendida apenas como aumento quantitativo da produtividade, mas como possibilidade de direcionamento da energia e do tempo humanos para atividades de maior valor agregado, como a análise crítica aprofundada da prova e de questões jurídicas complexas, a fundamentação de decisões e o atendimento mais humanizado aos jurisdicionados.

Por outro lado, a IAG oferece a possibilidade de reduzir inconsistências oriundas de vieses humanos ou desigualdades no tratamento de processos semelhantes. Por meio

do aprendizado contínuo e análise de grandes volumes de decisões judiciais, a IA pode auxiliar na identificação de padrões decisórios, possibilitando a uniformização de entendimentos e a redução de decisões contraditórias em casos análogos.

Esse benefício é particularmente relevante em um sistema judicial que busca maior previsibilidade e segurança jurídica. Ao identificar divergências interpretativas em casos similares, a IA tem o potencial de alertar magistrados sobre possíveis inconsistências, contribuindo para uma jurisprudência mais coerente e para a redução da litigiosidade decorrente de tratamentos desiguais.

Nessa mesma linha, existe a possibilidade de desenvolvimento de uma política de alinhamento ético na própria arquitetura dos sistemas de IAG, criando camadas de proteção que não só assegurem a conformidade das saídas com o ordenamento jurídico vigente, extirpando eventuais vieses discriminatórios, como também garantam a rastreabilidade das decisões e de suas fundamentações.

Diferentemente de sistemas puramente automatizados, essa modalidade de IAG pode ser programada para incorporar salvaguardas éticas desde sua concepção, incluindo verificações automáticas de conformidade constitucional, alertas sobre potenciais discriminações e mecanismos de auditoria que permitam o rastreamento completo do processo decisório.

Por fim, a implementação responsável da IAG no Judiciário pode funcionar como catalisador para processos mais amplos de modernização e inovação no Sistema de Justiça. A necessidade de desenvolver protocolos de uso ético, sistemas de supervisão e mecanismos de controle de qualidade para a IA pode impulsionar melhorias gerais nos processos de trabalho, na gestão de dados e na cultura organizacional dos tribunais.

A experiência acumulada com o uso supervisionado da IAG também pode gerar *insights* valiosos sobre como outras tecnologias emergentes podem ser incorporadas de forma responsável ao trabalho judicial, estabelecendo precedentes e metodologias para futuras inovações.

Paradoxalmente, um dos principais benefícios da IAG no contexto judicial pode ser o fortalecimento do protagonismo humano no processo decisório. Ao automatizar tarefas mais mecânicas e repetitivas, a IA libera os operadores do direito para se concentrarem naquilo que é genuinamente humano na atividade jurisdicional: a interpretação criativa do direito, a análise contextual dos casos, a ponderação de valores em conflito e a construção de soluções justas para situações únicas.

A premissa deve ser a de que a tecnologia deve assistir, e não dominar, a justiça (Beng, 2023), funcionando como um instrumento de apoio que otimiza o desempenho humano e permite que os operadores do Direito se concentrem nas tarefas que exigem pensamento crítico e análise aprofundada (Peixoto; Bonat, 2023).

## 7 CONCLUSÃO

Este artigo pretendeu reunir e sistematizar alguns conceitos e características da IA Generativa, de forma a favorecer sua melhor compreensão em termos de potencialidades, limitações e riscos. Teve também o objetivo de investigar em que medida, com adequado gerenciamento de riscos, essa inovação pode ser aproveitada para aperfeiçoar a atuação de magistrados e servidores e a atuação eficiente e efetiva do Poder Judiciário.

A pesquisa revelou que a hipótese inicialmente levantada é verdadeira. Os riscos da IAG no sistema de justiça, como uma via de mão dupla, residem em não saber aplicá-la, mas, também, paradoxalmente, em não fazer uso dela.

A implementação bem-sucedida da Inteligência Artificial Generativa no Poder Judiciário exige uma abordagem que transcenda a mera adoção tecnológica, demandando a construção de uma moldura robusta de governança pautada na ética e na supervisão humana. O desafio central reside em maximizar os benefícios transformadores da IAG, notadamente na gestão adequada do conhecimento disponível na instituição, enquanto se mitiga efetivamente seus riscos inerentes, preservando a integridade, a legitimidade e a confiabilidade do Sistema de Justiça.

Por fim, sem dúvidas, a resposta aos dilemas envolvendo Inteligência Artificial Generativa e Sistema de Justiça devem perpassar uma reflexão madura sobre inovação responsável e adequado protagonismo humano na inserção das novas tecnologias na atividade fim. Atribuir à máquina um poder que nunca foi seu, o de ditar a forma como as atividades devem ser feitas, talvez seja o primeiro erro do ser humano em sua interação com as novas tecnologias.

## REFERÊNCIAS

BAHNER, J. Elin; HÜPER, Anke-Dorothea; MANZEY, Dietrich. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. **International Journal of Human-Computer Studies**, London, v. 66, n. 9, p. 688-699, 2008.

BANH, Leonardo; STROBEL, Gero. Generative artificial intelligence. **Electronic Markets, Springer, IIM University of St. Gallen**, [*s. l.*], v. 33, n. 1, p. 1-17, 2023.

BENG, James Lau Oon. Generative Artificial Intelligence: Legal Profession Disrupted?. **International In-House Counsel Journal**, [*s. l.*], v. 16, n. 65, p. 8757-8764, Autumn 2023.

BEZERRA NETO, Bianor Arruda. Valoração da prova: o papel do juiz humano na era da inteligência artificial. **Conjur**, 8 mar. 2025. Disponível em https://www.conjur.com.br/2025-mar-08/valoracao-da-prova-o-papel-do-juiz-humano-na-era-da-inteligencia-artificial/. Acesso em: 5 set. 2025.

BONAT, Débora; VALE, Luís Manoel Borges do; PEREIRA, João Sergio dos Santos Soares. Inteligência artificial generativa e a fundamentação da decisão judicial. **Revista de Processo**, [*s. l.*], v. 346, p. 349-370, dez. 2023.

BRASIL. Conselho Nacional de Justiça. Resolução n.º 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. **Portal CNJ**, 11 mar. 2025. Disponível em: https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf. Acesso em: 20 mar. 2025.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. **Guia sobre Privacidade desde a Concepção e por Padrão**. Versão 1.0. Brasília, DF: MGI, 2024. Disponível em: https://www.gov.br/governodigital/pt-br/privacidade-e-seguranca/ppsi/guia_privacidade_concepcao.pdf. Acesso em: 7 set. 2025.

BUOLAMWINI, Joy A. **Gender shades**: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements of the degree of Master of Science at the Massachusetts Institute of Technology, September 2017.

CONSELHO NACIONAL DE JUSTIÇA - CNJ. **O uso da Inteligência Artificial generativa no Poder Judiciário brasileiro**: relatório de pesquisa. Brasília: CNJ, 2024.

DEEPSEEK: por que governos de países como Itália e Coreia do Sul barraram IA chinesa? **O Globo – Economia**, 8 fev. 2025. Disponível em: https://oglobo.globo.com/economia/noticia/2025/02/08/deepseek-por-que-governos-de-paises-como-italia-e-coreia-do-sul-barraram-ia-chinesa.ghtml?utm_source=chatgpt.com. Acesso em: 10 fev. 2025.

DIVINO, Sthéfano Bruno Santos. Simbiose tecnológica e do uso da Inteligência Artificial no Poder Judiciário: é preciso optar entre eficiência e segurança jurídica? **Boletim Revista dos Tribunais Online**, [*s. l.*], v. 63, maio 2025.

E-ESTONIA. **Artificial intelligence as the new reality of e-justice**. e-Estonia, 27 abr. 2020. Disponível em: https://e-estonia.com/artificial-intelligence-as-the-new-reality-of-e-justice/. Acesso em: 5 set. 2025.

FERRAZ, Taís S. A litigiosidade como fenômeno complexo: quanto mais se empurra, mais o sistema empurra de volta. **Revista Jurídica da Presidência**, Brasília, DF, v. 25, n. 135, p. 163-191, jan./abr. 2023. Disponível em: https://revistajuridica.presidencia.gov.br/index.php/saj/article/view/2847. Acesso em: 10 fev. 2025.

GABRIEL, Anderson de Paiva; PORTO, Fabio Ribeiro; ARAÚJO, Valter Shuenquener de. Justiça 4.0 e a Inteligência Artificial: pragmática eficiência. **Revista de Análise Econômica do Direito**, [*s. l.*], v. 9, 2025.

HATEM, Rami; SIMMONS, Brianna; THORBTON, Joseph E. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. **Cureus**, [*s. l.*], v. 15, n. 9, p. e44720, 2023.

HÖLLDOBLER, Steffen; MÖHLE, Sibylle; TIGUNOVA, Anna. **Lessons Learned from AlphaGo**. Disponível em https://ceur-ws.org/Vol-1837/paper14.pdf. Acesso em: 5 set. 2025.

JEVONS, William S. **The coal question; an inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines**. London: Macmillan and Co., 1865.

LIMÓN, Raúl. DeepSeek reports campaign of 'large-scale malicious attacks' after becoming most downloaded app. **El País**, 28 jan. 2025. Disponível em: https://english.elpais.com/technology/2025-01-28/deepseek-reports-campaign-of-large-scale-malicious-attacks-after-becoming-most-downloaded-app.html. Acesso em: 10 fev. 2025.

MICHAELSON, Greg. Programming Paradigms, Turing Completeness and Computational Thinking. **The Art, Science, and Engineering of Programming**, [*s. l.*], v. 4, n. 3, 2020. Disponível em: https://arxiv.org/pdf/2002.06178. Acesso em: 5 set. 2025.

MOLLICK, Ethan. **Co-Intelligence**: living and working with AI. New York: Penguim, 2024.

MÜNCH, Luciane, A. C.; FERRAZ, Taís S. Exploring defuturing to design Artificial-Intelligence artifacts: a systemic-design approach to tackle litigiousness in the brazilian judiciary. **Laws**, Switzerland, v. 13, n. 1, 2024. Disponível em: https://www.mdpi.com/2075-471X/13/1/4. Acesso em: 10 fev. 2025.

NAGLI, Gal. Wiz Research uncovers exposed DeepSeek database leaking sensitive information, including chat history. **Wiz**, 29 jan. 2025. Disponível em: https://www.wiz.io/blog/wiz-research-uncovers-exposed-deepseek-database-leak. Acesso em: 10 fev. 2025.

O'NEIL, Cathy. **Weapons of Math Destruction**: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Comitê para a Eliminação da Discriminação racial da Organização das Nações Unidas. **Recomendação n.º 26, de 24 de novembro de 2020**. Disponível em: https://acnudh.org/wp-content/uploads/2020/12/CERD_C_GC_36_PORT_REV.pdf. Acesso em: 10 fev. 2025.

PARASURAMAN, Raja; MANZEY, Dietrich. Complacency and bias in human use of automation: an attentional integration. **Human Factors: The Journal of the Human Factors and Ergonomics Society**, [*s. l.*], v. 52, n. 3, p. 381-410, 2010.

PEIXOTO, Fabiano Hartmann; BONAT, Debora. GPTs e Direito: impactos prováveis das IAs generativas nas atividades jurídicas brasileiras. Sequência: **Estudos Jurídicos e Políticos**, Florianópolis, v. 44, n. 93, 2023.

RODRIGUES, Leonel Cezar *et al*. Inteligência artificial, ética e celeridade no direito. Revista do CEJUR/TJSC: **Prestação Jurisdicional**, Florianópolis, v. 12, p. 1-19, 2024.

SANTOS, Adriano Cardoso dos. O uso da inteligência artificial pela suprema corte brasileira: desafios e potencialidades. **Revista Foco**, [*s. l.*], v. 18, n. 4, p. 1-23, 2025.

SENGE, Peter. **A quinta disciplina**: a arte e a prática da organização que aprende. Tradução Gabriel Zide Neto. Rio de janeiro: Best Seller, 2013.

STAR WARS: Episódio VI – O retorno do Jedi. Direção: Richard Marquand. Produção: George Lucas. Los Angeles: Lucasfilm, 1983.

THE CONVERSATION. OpenAI says DeepSeek inappropriately copied ChatGPT – but it's facing copyright claims too. **The Conversation**, 4 fev. 2025. Disponível em: https://theconversation.com/openai-says-deepseek-inappropriately-copied-chatgpt-but-its-facing-copyright-claims-too-248863. Acesso em: 10 fev. 2025.

THEOBALD, Oliver. **Machine learning for absolute beginners**: a plain english Introduction**.** 3th ed. New York: Jeremy Pederson and Red, 2017.

TURING, Alan. **Computing Machinery and Intelligence**. 1950. Disponível em: https://www.csee.umbc.edu/courses/471/papers/turing.pdf. Acesso em: 19 jun. 2019.

TURING, Alan. **On Computable Numbers, with an application to the Entscheidungsproblem**. 1936. Disponível em: http://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf. Acesso em: 19 jun. 2019.

WILKINS, Neil. **Artificial Intelligence**: what you need to know about machine learning, robotics, deep learning, recommender systems, Internet of Things, neural networks, reinforcement learning and our future. New York: Independently Published, 2019.

WORLD ECONOMIC FORUM. ChatWTO: An Analysis of Generative Artificial Intelligence and International Trade. **White Paper**. Geneva: World Economic Forum, 2024.

WORLD ECONOMIC FORUM. What is open-source AI and how could DeepSeek change the industry? **Weforum – Emerging Technologies**, 5 Feb. 2025. Disponível em: https://www.weforum.org/stories/2025/02/open-source-ai-innovation-deepseek/. Acesso em: 10 fev. 2025.

YIM, Hyunsu. South Korea spy agency says DeepSeek 'excessively' collects personal data. **Reuters**, 10 Feb. 2025. Disponível em: https://www.reuters.com/technology/artificial-

intelligence/south-korea-spy-agency-says-deepseek-excessively-collects-personal-data-2025-02-10/?utm_source=chatgpt.com. Acesso em: 10 fev. 2025.

## NOTA

Cíntia Menezes Brunetta e Taís Schilling Ferraz contribuíram de maneira equivalente para todas as etapas de concepção, estruturação e redação do presente artigo, bem como procederam à leitura e aprovação da versão final submetida para publicação. Alisson Carvalho de Alencar participou da redação dos trechos relativos aos riscos decorrentes do uso da inteligência artificial, notadamente no que se refere à proteção de dados pessoais, tendo igualmente realizado a leitura crítica integral do manuscrito e manifestado sua concordância quanto à versão final publicada.

**Como citar este documento:**

BRUNETTA, Cíntia Menezes; FERRAZ, Taís Schilling; ALENCAR, Alisson Carvalho de. Entre alucinações e inovações: o desafio da implementação responsável da IA Generativa no Sistema de Justiça. **Revista Opinião Jurídica**, Fortaleza, v. 23, n. 44, p. 1-26, set./dez. 2025. DOI: https://doi.org/10.12662/2447-6641oj.v23i43.p1-24.2025. Disponível em: link do artigo. Acesso em: xxxx.