ORGANIZADORES

Fabiane Regina Carvalho de Andrade Ibrahin
Fernando Martins Fontes

INTELIGÊNCIA ARTIFICIAL E SOCIEDADE

DESAFIOS ÉTICOS, JURÍDICOS E HUMANOS



MAX
EDITORA

Organizadores:

Fabiane Regina Carvalho de Andrade Ibrahin
Fernando Martins Fontes

INTELIGÊNCIA ARTIFICIAL E SOCIEDADE

DESAFIOS ÉTICOS, JURÍDICOS E HUMANOS

São Paulo
2026

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Inteligência artificial e sociedade [livro eletrônico] : desafios éticos, jurídicos e humanos / organizadores Fabiane Regina Carvalho de Andrade Ibrahin, Fernando Martins Fontes.
-- São Paulo : Max Editora, 2026.
PDF

Vários autores.
Bibliografia.
ISBN 978-65-86504-38-5

1. Ética 2. Inteligência artificial
3. Inteligência artificial - Aspectos sociais
4. Relações sociais 5. Tecnologia I. Ibrahin, Fabiane Regina Carvalho de Andrade.
II. Fontes, Fernando Martins.

26-416180.0 CDD-303.483

Índices para catálogo sistemático:

1. Inteligência artificial : Sociologia 303.483

Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

Capa
Matheus Ibrahin

Diagramação
Renata Peres

Max Editora
www.maxeditora.com.br

SUMÁRIO/ SUMMARY

APRESENTAÇÃO 05

ARTIGOS

Da inteligência artificial: entre a automação e a valorização humana 09
Fernando Fontes

Inteligência artificial e transformação das relações sociais: aspectos éticos, prevenção do plágio e diretrizes para o uso responsável.....27
Renata Job Borges de Figueiredo

Autonomia, vínculos e confiança nas interações entre pessoas e sistemas de inteligência artificial 45
Solange R. Aroeira

O princípio da proteção do trabalhador frente à inteligência artificial: limites ao poder dos algoritmos.....65
Fabiane Regina Carvalho de Andrade Ibrahin

Intelligences artificial e jurisdição: limites constitucionais ao uso de sistemas automatizados no poder judiciário brasileiro.....85
Dante Batista Silva

Inteligência artificial no poder judiciário: limites e impactos nas garantias processuais civis.....107
Eliana da Silva Araújo

Crimes cibernéticos e o golpe do falso advogado e binance 125
Mário Sérgio de Oliveira

APRESENTAÇÃO

A inteligência artificial ocupa posição cada vez mais relevante nas transformações que caracterizam a sociedade contemporânea. Sua crescente incorporação às atividades humanas, às relações sociais e às estruturas institucionais tem produzido mudanças significativas nas formas de comunicação, produção do conhecimento, organização do trabalho e exercício de direitos. Ao mesmo tempo que oferece novas possibilidades de inovação, desenvolvimento e aperfeiçoamento de processos, essa expansão tecnológica suscita questionamentos que ultrapassam os limites da técnica e alcançam dimensões fundamentais da experiência humana.

O avanço dos sistemas inteligentes, especialmente daqueles capazes de processar grandes volumes de dados, identificar padrões e participar de processos decisórios, exige uma reflexão crítica sobre os valores que devem orientar sua concepção, implementação e utilização. Questões relacionadas à transparência, à responsabilidade, à autonomia individual, à proteção de dados pessoais, à igualdade e à preservação dos direitos fundamentais assumem especial importância diante de tecnologias cujos efeitos se projetam sobre diferentes esferas da vida em sociedade.

Nesse contexto, a relação entre inteligência artificial e sociedade não pode ser compreendida exclusivamente sob a perspectiva da eficiência ou do progresso tecnológico. É necessário reconhecer que a inovação, embora constitua importante instrumento de transformação, não representa um fim em si mesma. Seu desenvolvimento deve estar associado à promoção do bem-estar coletivo, ao respeito à dignidade da pessoa humana e à construção de relações sociais mais justas, inclusivas e responsáveis.

É a partir dessa compreensão que se apresenta a obra coletiva *Inteligência Artificial e Sociedade: Desafios Éticos, Jurídicos e Humanos*, concebida como um espaço de reflexão acadêmica e diálogo interdisciplinar sobre as múltiplas repercussões da inteligência artificial na contemporaneidade.

A coletânea reúne diferentes perspectivas de análise que, embora desenvolvidas a partir de objetos e abordagens específicos, encontram um ponto de convergência na necessidade de compreender criticamente os impactos das transformações tecnológicas sobre os indivíduos, as instituições e a sociedade.

A diversidade temática constitui uma das características centrais desta publicação. As discussões transitam entre os fundamentos éticos da utilização da inteligência artificial, as transformações das relações sociais e humanas, os desafios decorrentes da automação e os limites jurídicos aplicáveis à incorporação de sistemas inteligentes em diferentes contextos. Também se inserem nesse debate questões relacionadas à proteção dos trabalhadores, às garantias processuais, à atividade jurisdicional e aos riscos associados à criminalidade no ambiente digital.

Essas diferentes abordagens evidenciam que os desafios impostos pela inteligência artificial não se restringem a uma única área do conhecimento. Ao contrário, exigem a aproximação entre saberes, metodologias e perspectivas capazes de contribuir para uma compreensão mais abrangente das consequências sociais, jurídicas e humanas da inovação tecnológica.

No campo jurídico, a expansão da inteligência artificial impõe a necessidade de examinar a adequação dos instrumentos normativos existentes e os limites de sua aplicação diante de novas formas de atuação tecnológica. A utilização de sistemas automatizados em atividades que envolvem direitos, deveres e interesses juridicamente protegidos exige especial atenção à preservação das garantias fundamentais, à transparência dos procedimentos e à possibilidade de responsabilização pelos efeitos das decisões produzidas ou influenciadas por essas tecnologias.

Sob a perspectiva ética, emergem questionamentos relacionados à legitimidade das escolhas incorporadas aos sistemas inteligentes, à reprodução de desigualdades, à prevenção de práticas discriminatórias e à necessidade de assegurar que os benefícios da inovação não sejam alcançados à custa da redução da autonomia ou da vulnerabilização de determinados grupos sociais.

Na dimensão humana, por sua vez, torna-se indispensável refletir sobre os efeitos da interação crescente entre pessoas e sistemas de inteligência artificial, especialmente no que se refere à confiança, à liberdade de escolha, às relações profissionais e à preservação de espaços de decisão e participação humana.

Essas questões demonstram que o debate contemporâneo sobre inteligência artificial não se limita à possibilidade de desenvolver tecnologias cada vez mais sofisticadas. Envolve, sobretudo, a responsabilidade de estabelecer parâmetros para que sua utilização permaneça compatível com os princípios que sustentam uma sociedade democrática e comprometida com a efetivação dos direitos fundamentais.

Nesse sentido, a produção científica desempenha papel essencial. A pesquisa acadêmica oferece instrumentos para identificar problemas, questionar pressupostos, examinar consequências e formular propostas que contribuam para o desenvolvimento de práticas tecnológicas mais transparentes, responsáveis e socialmente legítimas.

A presente obra insere-se nesse esforço coletivo de construção do conhecimento. Não pretende oferecer respostas definitivas a questões que permanecem em constante transformação, mas estimular o pensamento crítico, ampliar o diálogo entre diferentes campos científicos e contribuir para o aprofundamento de discussões que se mostram cada vez mais necessárias.

Os coordenadores desta publicação expressam seus sinceros agradecimentos aos autores que contribuíram para sua realização. O compromisso com a pesquisa, a dedicação intelectual e a diversidade das reflexões apresentadas conferem sentido a este projeto coletivo e reafirmam a importância da colaboração acadêmica na produção e na disseminação do conhecimento.

Estendem, igualmente, seu reconhecimento a todos que, de diferentes maneiras, participaram da concretização desta obra, contribuindo para que as reflexões aqui reunidas possam alcançar pesquisadores, docentes, estudantes, profissionais e demais interessados nos desafios contemporâneos da inteligência artificial.

Os coordenadores esperam que esta coletânea represente não apenas uma contribuição ao debate científico, mas também um convite à reflexão sobre o futuro que a sociedade deseja construir diante das transformações tecnológicas em curso.

Afinal, o verdadeiro desafio da inteligência artificial não reside apenas em ampliar as capacidades das máquinas, mas em assegurar que o desenvolvimento tecnológico permaneça orientado pela ética, pela responsabilidade e pelo respeito à dignidade humana.

Fabiane Regina Carvalho de Andrade Ibrahin
Fernando Martins Fontes

Da inteligência artificial: entre a automação e a valorização humana

Artificial intelligence: between automation and human valorization

*Fernando Fontes**

RESUMO

Este artigo analisa os impactos da inteligência artificial sobre a produtividade e o trabalho, com foco na automação de tarefas, na complementaridade entre pessoas e tecnologia e nas necessidades de qualificação profissional. Argumenta-se que os ganhos de produtividade não são automáticos nem uniformes: dependem de escolhas institucionais, políticas de capacitação, regulação ética e desenho organizacional que favoreçam a valorização do trabalho humano. Com base em literatura recente e relatórios de organismos internacionais, discute-se o risco de polarização ocupacional e as condições para que a IA amplie, em vez de restringir, as oportunidades de trabalho qualificado.

Palavras-chave: inteligência artificial; produtividade; automação; complementaridade; qualificação profissional.

ABSTRACT

This article analyzes the impacts of artificial intelligence on productivity and work, focusing on task automation, complementarity between people and technology, and professional qualification needs. It argues that productivity gains are neither automatic nor uniform: they depend on institutional choices, training policies, ethical regulation, and organizational design that favor the valorization of human work. Drawing on recent literature and reports from international organizations, the article discusses the risk of occupational polarization and the conditions under which AI can expand, rather than restrict, opportunities for skilled work.

Keywords: artificial intelligence; productivity; automation; complementarity; professional qualification.

* Professor há 15 anos. Diretor Acadêmico da Faculdade Lusófona em Cotia. Mais de 28 anos de atuação em empresas de grande porte, com experiência em alta gestão. MBA em Inteligência Artificial & Big Data. Pós-graduação em Gestão de Pessoas e Psicologia Organizacional. Pós-graduação em Gestão Ambiental. Mestre em Administração. MBA em Estratégia Empresarial. Graduação em Logística Empresarial. Graduação em Administração de Empresas.

1 INTRODUÇÃO

A inteligência artificial deixou de ser uma tecnologia emergente para tornar-se infraestrutura produtiva. Modelos generativos, sistemas de visão computacional, agentes autônomos e ferramentas de otimização já estão integrados a processos industriais, serviços, administração pública e trabalho do conhecimento. Essa difusão redefine a organização do trabalho, as competências demandadas e a distribuição dos ganhos de produtividade.

O debate sobre os efeitos da IA no emprego e na produtividade permanece aberto e polarizado. De um lado, estimativas de automatização de tarefas rotineiras e cognitivas geram preocupações com desemprego tecnológico e polarização ocupacional. De outro, evidências de complementaridade e de criação de novas ocupações sugerem que o resultado líquido depende menos da tecnologia em si e mais das instituições, das políticas de formação e das escolhas de desenho organizacional.

Este artigo examina essa tensão. Parte-se do reconhecimento de que a IA pode elevar a produtividade, mas que os benefícios só se materializam de forma ampla quando acompanhados de investimentos em capital humano, regulação adequada e estratégias que reforcem a colaboração entre pessoas e sistemas inteligentes. O texto está organizado em seis seções substantivas: a transformação da produtividade, a automação e a substituição de tarefas, a complementaridade, a qualificação profissional, os desafios éticos e as perspectivas futuras, seguidas das considerações finais.

A literatura recente, especialmente os relatórios da OECD (2023) e da International Labour Organization, indica que a adoção de IA já atinge setores tão diversos quanto manufatura avançada, serviços financeiros, saúde e educação. No entanto, a velocidade de difusão varia significativamente entre países e entre empresas de diferentes portes. Firmas líderes concentram a maior parte dos investimentos e dos ganhos iniciais, enquanto pequenas e médias empresas enfrentam barreiras de custo, de competências e de acesso a dados de qualidade. Essa heterogeneidade reforça a importância de políticas públicas que reduzam assimetrias e ampliem a capacidade de adoção produtiva.

Além disso, o debate contemporâneo sobre IA generativa introduziu novas dimensões. Diferentemente de ondas anteriores de automação, que afetavam principalmente tarefas rotineiras e manuais, os modelos

de linguagem e de imagem impactam atividades que exigem geração de conteúdo, síntese de informação e resolução de problemas não estruturados. Isso amplia tanto as oportunidades de complementaridade quanto os riscos de substituição em ocupações de média e alta qualificação. Compreender essas dinâmicas exige uma abordagem que combine evidências empíricas, análise institucional e atenção às condições de trabalho.

A dimensão setorial da adoção também merece atenção. Setores intensivos em conhecimento, como serviços financeiros, consultoria e desenvolvimento de software, registram taxas de experimentação e de implementação mais elevadas. Já atividades tradicionais de comércio, construção e serviços pessoais avançam de forma mais lenta, o que contribui para a heterogeneidade dos efeitos sobre o emprego e a produtividade agregada. Compreender essas diferenças é essencial para o desenho de políticas adequadas a cada contexto.

Finalmente, a literatura sobre mudança tecnológica enfatiza que os efeitos de curto prazo frequentemente diferem dos de médio e longo prazo. No curto prazo, os custos de adaptação, a incerteza e a concentração de benefícios podem dominar. No médio prazo, a difusão de competências complementares e a reorganização institucional tendem a ampliar os ganhos. Avaliar a trajetória da IA exige, portanto, horizontes temporais que ultrapassem os ciclos imediatos de adoção.

O presente artigo adota uma perspectiva que combina contribuições da economia do trabalho, da sociologia das organizações e dos estudos de ciência, tecnologia e sociedade. Essa abordagem interdisciplinar permite examinar tanto os efeitos mensuráveis sobre produtividade e emprego quanto as dimensões qualitativas relacionadas à autonomia, ao sentido do trabalho e à distribuição de poder no interior das organizações. A análise se apoia em literatura acadêmica recente e em relatórios de organismos internacionais, sem pretender esgotar a diversidade de experiências nacionais e setoriais.

Ao longo do texto, busca-se evitar tanto o otimismo tecnológico irrestrito quanto o pessimismo que antecipa desemprego massivo. O argumento central é que os resultados dependem de mediações institucionais e organizacionais que podem ser objeto de intervenção deliberada. Identificar essas mediações e as condições sob as quais elas favorecem a valorização do trabalho humano constitui o objetivo principal das seções que se seguem.

2 A INTELIGÊNCIA ARTIFICIAL E A TRANSFORMAÇÃO DA PRODUTIVIDADE

Os ganhos de produtividade associados à IA são heterogêneos e ainda em consolidação. Em setores intensivos em dados e em tarefas padronizáveis — como manufatura avançada, logística, serviços financeiros e desenvolvimento de software — observa-se elevação mensurável da eficiência. Relatórios da OCDE e do Fórum Econômico Mundial indicam que empresas que combinam adoção de IA com redesenho de processos e capacitação de equipes registram aumentos de produtividade superiores àquelas que apenas implementam a tecnologia de forma isolada.

Contudo, a literatura econômica alerta para o chamado "paradoxo da produtividade": a difusão tecnológica nem sempre se traduz imediatamente em crescimento agregado. Fatores como custos de adaptação, incerteza regulatória, escassez de competências complementares e concentração de benefícios em firmas líderes retardam os ganhos macroeconômicos. Além disso, a IA generativa introduz uma novidade: ela afeta não apenas tarefas rotineiras, mas também atividades que exigem geração de conteúdo, análise e resolução de problemas não estruturados.

A transformação em curso, portanto, não se resume à substituição de trabalho por capital. Ela reconfigura a fronteira entre tarefas automatizáveis e tarefas que se beneficiam da combinação de julgamento humano e capacidade computacional. O resultado sobre a produtividade agregada dependerá da velocidade com que organizações e sistemas educacionais se ajustarem a essa reconfiguração.

Estudos de caso em indústrias de manufatura avançada mostram que a integração de sistemas de visão computacional e de manutenção preditiva pode reduzir tempos de parada e aumentar a taxa de conformidade de qualidade. Contudo, os ganhos dependem da capacidade das equipes de interpretar os alertas gerados pelos modelos e de reorganizar fluxos de trabalho. Quando essa reorganização não ocorre, a tecnologia permanece subutilizada e os retornos ficam abaixo das expectativas iniciais.

No setor de serviços, ferramentas de atendimento automatizado e de análise de documentos têm permitido que profissionais dediquem mais tempo a interações de maior valor. Ao mesmo tempo, a literatura alerta para o risco de que a pressão por volume de atendimento se intensifique, transferindo carga cognitiva para os trabalhadores que supervisionam os

sistemas. A produtividade medida por indicadores tradicionais pode crescer, enquanto a qualidade da experiência de trabalho e a sustentabilidade do desempenho no médio prazo se deterioram.

Do ponto de vista macroeconômico, a velocidade com que os ganhos de produtividade se difundem para o conjunto da economia depende de fatores institucionais. Mercados de trabalho flexíveis, sistemas de educação e formação profissional responsivos e políticas de concorrência que evitem concentração excessiva de poder de mercado são condições frequentemente associadas a uma difusão mais ampla dos benefícios. Sem essas condições, a IA tende a ampliar disparidades entre firmas e entre trabalhadores.

A medição da produtividade no contexto da IA apresenta desafios metodológicos. Indicadores tradicionais de produtividade do trabalho podem não capturar adequadamente melhorias de qualidade, redução de erros ou ganhos de tempo que não se traduzem imediatamente em aumento de volume de produção. Além disso, parte dos benefícios pode se materializar na forma de novos produtos e serviços, cuja contribuição para o valor agregado é de difícil mensuração nos primeiros estágios de difusão.

Empresas que combinam investimentos em IA com reorganização de processos e com formação de equipes multidisciplinares tendem a reportar retornos mais elevados. A complementaridade entre capital digital e capital organizacional aparece de forma consistente em estudos de caso e em análises econométricas. Isso sugere que a tecnologia, isoladamente, não constitui o principal determinante dos ganhos; o modo como ela é incorporada às práticas de trabalho é decisivo.

A difusão desigual da IA também se reflete em diferenças de produtividade entre firmas do mesmo setor. Empresas que conseguem integrar os sistemas aos seus processos específicos e que dispõem de dados de qualidade tendem a se distanciar das demais. Esse fenômeno de divergência de produtividade tem sido documentado em vários países e reforça a importância de políticas que facilitem o acesso de pequenas e médias empresas a ferramentas, a formação e a infraestrutura de dados.

Além dos efeitos diretos sobre a eficiência operacional, a IA pode contribuir para a inovação de produtos e de modelos de negócio. Ferramentas de geração e de análise permitem explorar combinações de características e antecipar demandas com maior rapidez. Os ganhos, contudo, dependem da capacidade das organizações de traduzir insights em ofertas comercializáveis e de gerir os riscos associados a novos produtos. A produtividade de longo prazo, portanto, incorpora tanto eficiência quanto capacidade inovativa.

3 AUTOMAÇÃO E SUBSTITUIÇÃO DE TAREFAS

A automação impulsionada pela inteligência artificial concentra-se, em primeira instância, em tarefas rotineiras e previsíveis. Estudos clássicos de Frey e Osborne, bem como análises mais recentes da OIT e da OCDE, estimam que uma parcela significativa das ocupações atuais contém tarefas altamente expostas à automatização. No entanto, a substituição completa de postos de trabalho permanece limitada: a maioria das ocupações combina tarefas automatizáveis com outras que exigem interação social, julgamento contextual, criatividade ou responsabilidade.

A IA generativa ampliou o alcance da automação para atividades antes consideradas cognitivas e não rotineiras redação, programação, análise de documentos, atendimento e design. Isso não implica, necessariamente, eliminação massiva de empregos. Em muitos casos, observa-se a reorganização do conteúdo das tarefas: o trabalhador passa a supervisionar, revisar, contextualizar e aplicar o output dos sistemas, em vez de executar todas as etapas manualmente.

O risco mais concreto, portanto, não é o desemprego tecnológico generalizado no curto prazo, mas a polarização. Ocupações de média qualificação, compostas por tarefas parcialmente automatizáveis, tendem a se contrair ou a se fragmentar, enquanto crescem tanto postos de alta qualificação (que combinam IA e expertise) quanto atividades de baixa remuneração e baixa automatização. A literatura de Autor sobre polarização ocupacional continua relevante e encontra novos contornos com a difusão da IA generativa.

Empresas e políticas públicas enfrentam o desafio de gerir essa transição. A ausência de estratégias de realocação e de proteção de renda pode concentrar os custos da automação sobre trabalhadores específicos, mesmo quando a produtividade agregada aumenta.

A análise de tarefas, proposta por Autor e outros autores, oferece um quadro útil para compreender os limites da substituição. Tarefas que envolvem reconhecimento de padrões em dados estruturados, processamento de linguagem em contextos previsíveis e operação de equipamentos em ambientes controlados são mais suscetíveis à automação. Já atividades que exigem interação social complexa, julgamento ético em situações ambíguas ou adaptação a ambientes físicos não estruturados permanecem, por ora, mais resistentes.

Evidências do mercado de trabalho brasileiro e de outros países emergentes indicam que a polarização ocupacional não se manifesta de forma idêntica à observada em economias avançadas. A elevada informalidade e a baixa intensidade tecnológica de parte expressiva do emprego formal limitam tanto o ritmo de substituição quanto a capacidade de criação de postos de alta qualificação. Políticas de formalização e de elevação da base tecnológica das empresas tornam-se, nesse contexto, condições prévias para que os efeitos da IA se distribuam de modo mais equilibrado.

Além da substituição direta, a automação pode induzir a fragmentação de ocupações. Atividades que antes eram realizadas por um mesmo profissional passam a ser decompostas em subtarefas, algumas das quais são automatizadas e outras reorganizadas. Esse processo pode gerar tanto ganhos de eficiência quanto perda de autonomia e de sentido do trabalho, especialmente quando os trabalhadores perdem o controle sobre o conjunto da atividade e passam a executar apenas as partes residuais.

A experiência de setores como call centers e processamento de documentos mostram que a introdução de IA generativa pode reduzir o tempo de execução de tarefas, mas também intensificar o monitoramento e a padronização do desempenho. Sem mecanismos de participação dos trabalhadores no redesenho dos processos, o resultado tende a ser a compressão de salários relativos e o aumento da rotatividade, mesmo quando a produtividade agregada da empresa cresce.

A velocidade de substituição também depende da elasticidade da demanda pelos produtos e serviços afetados. Quando a automação reduz custos e preços, a demanda pode se expandir o suficiente para compensar, total ou parcialmente, a redução de emprego por unidade de produto. Em mercados saturados ou com baixa elasticidade, o efeito líquido sobre o emprego tende a ser mais negativo. Essa dinâmica ajuda a explicar por que os impactos observados variam entre setores e ao longo do tempo.

Trabalhadores em ocupações de média qualificação enfrentam, em muitos casos, a combinação de risco de automação parcial e de dificuldade de transição para funções de maior complexidade. A ausência de credenciais formais reconhecidas e de redes de apoio profissional pode prolongar períodos de desemprego ou de subemprego. Políticas de intermediação de mão de obra e de reconhecimento de competências adquiridas no trabalho tornam-se relevantes nesse cenário.

A recomposição de tarefas dentro das ocupações também altera as exigências de qualificação de forma não linear. Algumas funções passam a exigir maior capacidade de supervisão e de resolução de exceções, enquanto outras se tornam mais rotineiras e monitoráveis. Essa reconfiguração interna das ocupações dificulta previsões simples sobre o volume de emprego e reforça a necessidade de análises detalhadas por tarefa e por setor.

Em economias com elevada informalidade, a automação formal pode coexistir com a persistência de formas de trabalho pouco protegidas. Trabalhadores informais frequentemente não têm acesso às ferramentas nem às oportunidades de requalificação, o que pode ampliar a distância entre segmentos do mercado de trabalho. Políticas que articulem formalização, acesso a tecnologia e formação tornam-se particularmente relevantes nesses contextos.

4 COMPLEMENTARIDADE ENTRE PESSOAS E TECNOLOGIA

A visão de complementaridade enfatiza que a inteligência artificial, em vez de substituir o trabalho humano de forma abrangente, pode ampliar suas capacidades. Quando bem integrada, a tecnologia assume tarefas de processamento intensivo de dados, geração de opções e execução repetitiva, liberando as pessoas para atividades de maior valor: definição de objetivos, julgamento ético, negociação, inovação e cuidado.

Evidências de adoção empresarial mostram que os maiores ganhos de produtividade ocorrem em organizações que redesenham processos e capacitam equipes para trabalhar *com* os sistemas, e não apenas *ao lado* deles. Modelos de "copiloto" em que a IA sugere, o humano decide e ambos iteram têm se mostrado mais eficazes do que tentativas de automação total em contextos complexos e ambíguos.

Essa complementaridade, contudo, não é automática. Ela exige investimentos em interfaces adequadas, transparência dos sistemas, treinamento específico e culturas organizacionais que valorizem o julgamento humano. Sem essas condições, a tecnologia pode ser subutilizada ou gerar sobrecarga cognitiva e perda de autonomia.

Do ponto de vista do mercado de trabalho, a complementaridade favorece a criação de novas ocupações e a elevação da produtividade de trabalhadores já qualificados. Ao mesmo tempo, reforça a demanda por competências socioemocionais, pensamento crítico e capacidade de colaboração com sistemas inteligentes habilidades que se tornam diferenciais competitivos.

A complementaridade efetiva exige mais do que a simples disponibilização de ferramentas. Ela depende do desenho de interfaces que tornem o raciocínio dos modelos compreensível para o usuário humano, da definição clara de responsabilidades e da criação de espaços de deliberação em que o julgamento profissional possa ser exercido. Quando essas condições estão ausentes, a tecnologia tende a ser utilizada de forma mecânica ou a gerar sobrecarga de verificação.

Experiências em ambientes de saúde e de educação ilustram o potencial e os limites da complementaridade. Sistemas de apoio à decisão clínica podem ampliar a capacidade diagnóstica de profissionais, desde que os alertas sejam contextualizados e que o médico conserve autoridade sobre a conduta. De modo análogo, ferramentas de geração de conteúdo educacional podem apoiar docentes na preparação de materiais, mas não substituem a mediação pedagógica e a adaptação às necessidades específicas dos estudantes.

Do ponto de vista organizacional, a complementaridade se fortalece quando as empresas investem simultaneamente em tecnologia e em práticas de gestão do trabalho. Programas de formação contínua, canais de feedback dos usuários e participação de equipes no redesenho de processos aumentam a probabilidade de que a IA seja integrada de modo a elevar tanto a produtividade quanto a qualidade do emprego. A literatura sobre mudança tecnológica enfatiza que os retornos dependem tanto do capital físico e digital quanto do capital organizacional.

Em contextos de alta incerteza regulatória, a complementaridade também envolve a construção de confiança. Trabalhadores tendem a colaborar mais efetivamente com sistemas cujos limites e vieses são conhecidos e cujas decisões podem ser contestadas. Transparência algorítmica e mecanismos de revisão humana não são apenas exigências éticas; constituem condições práticas para que a tecnologia amplie, em vez de restringir, a capacidade de julgamento das pessoas.

A complementaridade também se manifesta na criação de novas ocupações que não existiam antes da difusão da IA. Funções de curadoria de dados, de avaliação de vieses, de integração de sistemas e de mediação entre modelos e usuários finais têm surgido em empresas que adotam a tecnologia de forma estratégica. Essas ocupações exigem combinações específicas de conhecimento técnico e de compreensão do domínio de aplicação, reforçando a importância de trajetórias formativas flexíveis.

Em ambientes de trabalho colaborativo, a qualidade da complementaridade depende da clareza com que as equipes definem o que permanece sob responsabilidade humana e o que pode ser delegado aos sistemas. Protocolos explícitos de revisão, de escalonamento de decisões e de registro de intervenções humanas aumentam a confiabilidade dos processos e reduzem o risco de erros silenciosos. A ausência desses protocolos pode gerar tanto subutilização quanto dependência excessiva.

A complementaridade bem-sucedida também depende da qualidade dos dados e da adequação dos modelos aos contextos de uso. Sistemas treinados em dados que não representam a diversidade de situações reais podem gerar recomendações inadequadas ou viesadas, reduzindo a confiança dos usuários e limitando a adoção efetiva. Investimentos em curadoria de dados e em avaliação contínua de desempenho dos modelos constituem, portanto, parte integrante de estratégias de complementaridade.

Do ponto de vista dos trabalhadores, a percepção de que a tecnologia amplia sua capacidade, em vez de ameaçar sua posição, influencia a disposição para colaborar e para aprender. Quando a implementação é acompanhada de comunicação clara sobre objetivos, de garantia de que ganhos de produtividade não se traduzirão automaticamente em redução de quadro e de oportunidades de desenvolvimento, a resistência tende a diminuir e os resultados melhoram.

5 QUALIFICAÇÃO PROFISSIONAL

A transformação produtiva impulsionada pela IA eleva a importância da qualificação contínua. Relatórios do Fórum Econômico Mundial e da OCDE destacam que a maioria das competências demandadas até o final da década combina domínio técnico (incluindo uso de ferramentas de IA), pensamento analítico e habilidades humanas. A obsolescência de conhecimentos específicos acelera, enquanto a capacidade de aprender e de se adaptar ganha valor estratégico.

Sistemas de educação e formação profissional enfrentam o desafio de responder em velocidade compatível com a mudança tecnológica. Isso inclui a integração de conteúdos sobre IA nos currículos, a expansão de microcredenciais e formações flexíveis, e o fortalecimento de políticas de aprendizado ao longo da vida. Experiências bem-sucedidas combinam formação técnica com desenvolvimento de competências transversais.

No âmbito empresarial, programas de requalificação interna têm se mostrado mais eficazes do que a simples substituição de trabalhadores. Empresas que investem em capacitação junto com a adoção de IA registram maiores retornos e menor rotatividade. Do ponto de vista público, políticas de transição justa que associem proteção de renda, orientação profissional e formação reduzem os custos sociais da mudança tecnológica.

A qualificação, portanto, não é apenas uma resposta individual ao risco de automação. É condição estrutural para que os ganhos de produtividade se distribuam de forma mais ampla e para que a complementaridade entre pessoas e tecnologia se realize na prática.

A qualificação profissional no contexto da IA não se resume à aquisição de competências técnicas de programação ou de ciência de dados. Embora essas habilidades sejam relevantes em determinadas funções, a maioria dos trabalhadores precisa desenvolver capacidades de uso crítico de ferramentas, de interpretação de resultados e de colaboração com sistemas inteligentes. Competências socioemocionais, pensamento sistêmico e capacidade de aprendizagem contínua tornam-se diferenciais transversais.

Sistemas de educação e formação profissional enfrentam o desafio de atualizar currículos em ritmo compatível com a velocidade da mudança tecnológica. A literatura internacional aponta para a importância de abordagens modulares, de certificação de competências específicas e de parcerias entre instituições de ensino e empresas. No Brasil, a baixa densidade de formação técnica e a desigualdade de acesso a educação de qualidade amplificam o risco de que os benefícios da IA se concentrem em uma parcela restrita da força de trabalho.

Programas de requalificação interna têm se mostrado mais eficazes quando combinam formação em ferramentas específicas com redesenho de funções e apoio à transição de carreira. Empresas que tratam a capacitação como investimento estratégico, e não como custo residual, registram menores taxas de rotatividade e maior aproveitamento das tecnologias implantadas. Do lado público, políticas de transição justa que articulem proteção de renda, orientação profissional e formação podem reduzir os custos sociais da mudança.

A dimensão territorial também importa. Regiões com menor densidade de instituições de ensino superior e de centros de formação técnica tendem a enfrentar maiores dificuldades para acompanhar a demanda por novas competências. Políticas de desenvolvimento regional que incorporem infraestrutura digital e oferta educacional descentralizada constituem condições para que a qualificação profissional não aprofunde desigualdades espaciais já existentes.

A formação de formadores constitui um gargalo frequentemente subestimado. Docentes e instrutores de educação profissional precisam eles próprios desenvolver familiaridade com as ferramentas e com as implicações pedagógicas da IA. Sem essa atualização, a oferta de capacitação permanece desalinhada das demandas reais do mercado de trabalho. Investimentos em desenvolvimento docente e em recursos educacionais abertos adaptados a contextos locais são condições para ampliar a escala da qualificação.

Experiências de aprendizagem no local de trabalho, apoiadas por tutoria e por projetos práticos, têm se mostrado particularmente eficazes para a aquisição de competências de uso de IA. Diferentemente de cursos genéricos, essas abordagens permitem que os trabalhadores relacionem as ferramentas às tarefas concretas que realizam, aumentando a retenção e a aplicação do conhecimento. Empresas que estruturam esses percursos registram maiores taxas de adoção efetiva das tecnologias implantadas.

A qualificação também envolve a capacidade de avaliar criticamente as saídas dos sistemas de IA. Trabalhadores precisam desenvolver critérios para identificar alucinações, vieses e limitações de contexto, especialmente em aplicações de alto risco. Essa competência de literacia algorítmica não se adquire apenas por meio de cursos técnicos; exige prática supervisionada e discussão de casos reais. Instituições de formação que incorporam essa dimensão crítica preparam melhor os profissionais para ambientes de trabalho híbridos.

Finalmente, a sustentabilidade das estratégias de qualificação depende de sua articulação com políticas de proteção social. Períodos de transição entre ocupações podem exigir apoio de renda e de orientação. Sem essa rede de segurança, trabalhadores em situação de maior vulnerabilidade tendem a aceitar empregos de menor qualidade ou a permanecer em funções em declínio, reduzindo a mobilidade necessária para a adaptação produtiva.

A construção de ecossistemas de aprendizagem ao longo da vida, que articulem educação formal, formação profissional e aprendizagem no trabalho, constitui um dos principais desafios institucionais da era da inteligência artificial. Países que conseguirem organizar essas trajetórias de forma flexível e acessível estarão em melhor posição para transformar a mudança tecnológica em oportunidade de elevação da qualidade do emprego e da produtividade agregada. A ausência de tais ecossistemas, por outro lado, tende a perpetuar segmentações e a limitar o potencial inclusivo da tecnologia.

6 DESAFIOS ÉTICOS E PERSPECTIVAS FUTURAS

A difusão da inteligência artificial no trabalho levanta questões éticas e de governança que ultrapassam a dimensão puramente econômica. Entre os principais desafios estão a transparência e a explicabilidade dos sistemas, o risco de vieses algorítmicos na seleção e avaliação de pessoas, a proteção de dados e privacidade, e a responsabilidade por decisões tomadas com apoio de IA. A concentração de poder em poucas empresas desenvolvedoras e a assimetria de informação entre empregadores e trabalhadores também demandam atenção regulatória.

Perspectivas futuras apontam para maior integração de agentes autônomos e sistemas multimodais, o que pode intensificar tanto a automação quanto as possibilidades de complementaridade. Relatórios recentes do Fórum Econômico Mundial e da OCDE projetam crescimento líquido de empregos em alguns cenários, desde que acompanhados de políticas ativas de formação e de regulação que incentive a inovação responsável. Ao mesmo tempo, persistem incertezas sobre a velocidade de adoção, os efeitos distributivos e a capacidade institucional de resposta.

Uma governança eficaz combina regulação setorial, padrões de transparência, participação de trabalhadores no desenho de sistemas e investimentos públicos em infraestrutura digital e capacitação. Sem esses elementos, os ganhos de produtividade tendem a se concentrar, ampliando desigualdades. Com eles, a IA pode contribuir para elevar a qualidade do trabalho e a produtividade de forma mais inclusiva.

A concentração de poder em poucas empresas desenvolvedoras de modelos de base levanta questões de soberania tecnológica e de dependência. Países e organizações que não participam do desenvolvimento e do treinamento desses modelos enfrentam limites para adaptar as ferramentas a contextos locais, para auditar vieses e para garantir que os dados utilizados respeitem normas de proteção. A construção de capacidades nacionais e regionais de desenvolvimento e de avaliação de sistemas torna-se, portanto, elemento de estratégia de desenvolvimento.

A proteção de dados e a privacidade adquirem contornos específicos quando sistemas de IA são utilizados para monitoramento de desempenho, seleção de candidatos ou avaliação contínua de trabalhadores. A coleta massiva de dados comportamentais pode gerar assimetrias de informação e restringir a autonomia individual. Regulamentações que estabeleçam limites

claros para o uso de dados no ambiente de trabalho, direito de explicação e mecanismos de contestação constituem salvaguardas necessárias para que a adoção tecnológica não eroda direitos fundamentais.

Perspectivas futuras incluem o desenvolvimento de sistemas mais transparentes e auditáveis, a expansão de modelos de código aberto e o fortalecimento de instituições de governança multissetorial. A participação de trabalhadores e de suas representações no desenho e na avaliação de sistemas de IA no local de trabalho tem sido apontada como fator que aumenta tanto a legitimidade quanto a eficácia das implementações. Sem essa participação, o risco de resistência e de subutilização permanece elevado.

No horizonte de médio prazo, a integração de IA com outras tecnologias como robótica avançada, internet das coisas e computação de borda tende a ampliar o escopo de tarefas potencialmente automatizáveis. Ao mesmo tempo, abre espaço para novas formas de complementaridade em ambientes físicos complexos. O resultado líquido continuará a depender das escolhas institucionais e organizacionais que moldam a direção da mudança tecnológica.

A governança da IA no trabalho também envolve questões de representação e de voz. Trabalhadores e suas organizações precisam de canais institucionais para participar da definição de regras de uso, de critérios de avaliação de desempenho assistido por algoritmos e de mecanismos de reparação. Experiências de negociação coletiva que incorporam cláusulas sobre tecnologias digitais mostram que é possível estabelecer salvaguardas sem impedir a inovação.

No plano internacional, a divergência de abordagens regulatórias entre jurisdições cria incertezas para empresas que operam em múltiplos países. A busca por padrões mínimos comuns, especialmente em temas de transparência, de responsabilidade e de proteção de dados, pode reduzir custos de conformidade e evitar corridas regulatórias para o fundo. Ao mesmo tempo, a preservação de espaço para adaptação a contextos nacionais permanece necessária.

A questão da responsabilidade por decisões apoiadas por IA permanece em aberto em muitos ordenamentos jurídicos. Definir quem responde por erros o desenvolvedor, a empresa usuária ou o profissional que validou a recomendação é condição para a segurança jurídica e para a confiança dos usuários. Regulamentações que estabeleçam deveres de diligência e de documentação das intervenções humanas contribuem para esclarecer essas responsabilidades sem impedir a inovação.

Perspectivas de médio prazo incluem o avanço de técnicas de IA explicável e de sistemas que permitam maior controle por parte dos usuários finais. Esses desenvolvimentos podem reduzir a opacidade e facilitar a complementaridade. Contudo, sua adoção efetiva dependerá de incentivos de mercado e de exigências regulatórias que valorizem a transparência tanto quanto o desempenho bruto dos modelos.

7 CONSIDERAÇÕES FINAIS

Os impactos da inteligência artificial sobre a produtividade e o trabalho não estão predeterminados pela tecnologia. Eles resultam de escolhas institucionais, organizacionais e políticas. A automação de tarefas é real e continuará avançando, mas a substituição completa de ocupações permanece limitada. O resultado mais provável, no médio prazo, é a reconfiguração do conteúdo do trabalho e a polarização, a menos que sejam adotadas estratégias deliberadas de complementaridade e qualificação.

Este artigo sustenta que os ganhos de produtividade só se traduzem em benefícios amplos quando a adoção tecnológica é acompanhada de redesenho de processos, investimentos em capital humano e regulação ética. A valorização do trabalho humano não é incompatível com a automação; ao contrário, pode ser reforçada por ela quando as pessoas passam a exercer funções de julgamento, supervisão e criação em colaboração com sistemas inteligentes.

A transição em curso exige ação coordenada de empresas, sistemas educacionais e políticas públicas. Sem essa articulação, os riscos de exclusão e desigualdade aumentam. Com ela, a era da inteligência artificial pode ampliar tanto a produtividade quanto as oportunidades de trabalho qualificado e digno.

A transição em curso não é determinada pela tecnologia de forma unilateral. Ela resulta da interação entre capacidades técnicas, arranjos institucionais, políticas de formação e escolhas de desenho organizacional. Reconhecer essa contingência é condição para evitar tanto o determinismo tecnológico quanto a subestimação dos riscos de exclusão e de polarização.

Empresas que tratam a adoção de IA como processo de transformação do trabalho, e não apenas como substituição de tarefas, tendem a obter retornos mais sustentáveis. Isso implica investir em interfaces compreensíveis, em formação contínua e em práticas de gestão que

preservem e valorizem o julgamento humano. Do lado público, a articulação entre regulação ética, proteção social e políticas de qualificação constitui o núcleo de uma estratégia de transição justa.

O desafio central, portanto, não é frear a difusão da inteligência artificial, mas orientar sua trajetória de modo que os ganhos de produtividade se distribuam de forma mais ampla e que as condições de trabalho se elevem. Isso exige coordenação entre atores privados, sistemas educacionais e políticas públicas, bem como atenção permanente aos efeitos distributivos e às dimensões éticas da mudança tecnológica.

Em síntese, a valorização do trabalho humano não é incompatível com a automação. Ao contrário, constitui condição para que a inteligência artificial contribua de modo efetivo para o aumento da produtividade e para a melhoria das oportunidades de emprego qualificado. As escolhas feitas nos próximos anos em empresas, em instituições de ensino e em arenas regulatórias definirão se essa complementaridade se realizará na prática ou permanecerá como possibilidade não realizada.

As evidências disponíveis até o momento não autorizam conclusões definitivas sobre o saldo líquido de emprego no longo prazo. O que se observa com maior clareza é a reconfiguração das tarefas dentro das ocupações e a elevação da demanda por competências complementares. Nesse cenário, a qualidade das instituições de formação e de proteção social torna-se determinante para que a mudança tecnológica não se traduza em exclusão duradoura de grupos específicos de trabalhadores.

Por fim, a valorização do trabalho humano no contexto da inteligência artificial exige que as métricas de sucesso das organizações e das políticas públicas incorporem não apenas ganhos de produtividade, mas também indicadores de qualidade do emprego, de autonomia e de distribuição dos benefícios. Sem essa ampliação do horizonte de avaliação, o risco de que a tecnologia avance de forma desconectada das condições de vida e de trabalho permanece elevado.

As experiências internacionais sugerem que países que combinaram investimentos em infraestrutura digital, em educação e em proteção social conseguiram capturar uma parcela maior dos benefícios da digitalização anterior. A lição permanece válida para a onda atual de inteligência artificial. A ausência de coordenação entre essas dimensões tende a concentrar ganhos e a externalizar custos sociais, reduzindo o potencial de crescimento inclusivo.

Em conclusão, o artigo sustenta que a inteligência artificial pode contribuir para a elevação da produtividade e para a criação de oportunidades de trabalho qualificado, desde que acompanhada de escolhas institucionais e organizacionais que valorizem o trabalho humano. A automação e a complementaridade não são trajetórias mutuamente exclusivas; sua combinação depende de políticas de formação, de regulação ética e de práticas de gestão que coloquem as pessoas no centro do processo de transformação tecnológica.

REFERÊNCIAS BIBLIOGRÁFICAS

BRYNJOLFSSON, Erik; MCAFEE, Andrew. **The Second Machine Age**: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W. W. Norton & Company, 2014.

David H. Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, v. 29, n. 3, p. 3-30, 2015.

FREY, Carl Benedikt; OSBORNE, Michael A. **The future of employment**: how susceptible are jobs to computerisation? Technological Forecasting and Social Change, v. 114, p. 254-280, 2017.

INTERNATIONAL LABOUR ORGANIZATION (ILO). **Generative AI and Jobs**: A Global Analysis of Potential Effects on Job Quantity and Quality. Geneva: ILO, 2023.

OECD. OECD. **Employment Outlook 2023**: Artificial Intelligence and the Labour Market. Paris: OECD Publishing, 2023.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: A Modern Approach. 4. ed. Hoboken: Pearson, 2021.

SCHWAB, Klaus. **A Quarta Revolução Industrial**. São Paulo: Edipro, 2016.

WORLD ECONOMIC FORUM. **The Future of Jobs Report 2025**. Geneva: WEF, 2025.; https://www.academia.edu/180273619/DA_INTELIG%C3%8ANCIA_ARTIFICIAL_ENTRE_A_AUTOMA%C3%87%C3%83O_E_A_VALORIZA%C3%87%C3%83O_HUMANA

INTELIGÊNCIA ARTIFICIAL E TRANSFORMAÇÃO DAS RELAÇÕES SOCIAIS: ASPECTOS ÉTICOS, PREVENÇÃO DO PLÁGIO E DIRETRIZES PARA O USO RESPONSÁVEL

ARTIFICIAL INTELLIGENCE AND THE TRANSFORMATION OF SOCIAL RELATIONS: ETHICAL ASPECTS, PLAGIARISM PREVENTION, AND GUIDELINES FOR RESPONSIBLE USE

*Renata Job Borges de Figueiredo**

RESUMO

Este artigo analisa como a inteligência artificial (IA) generativa transforma as relações sociais, educacionais e informacionais, com ênfase nos dilemas éticos, na integridade acadêmica e na prevenção do plágio. Por meio de revisão bibliográfica narrativa, discutem-se os princípios éticos consolidados na literatura, os riscos de viés, opacidade e desinformação, e os limites dos detectores automáticos de texto. Propõem-se diretrizes para o uso responsável da IA em contextos acadêmicos, fundamentadas em transparência, autoria humana, verificação e equidade. Conclui-se que a prevenção do plágio exige menos vigilância tecnológica e mais educação para a integridade, redesenho avaliativo e governança institucional clara.

Palavras-chave: inteligência artificial; ética; plágio; integridade acadêmica; relações sociais; uso responsável.

---

* Especialista em Política Públicas, Filosofia do Direito, Metodologia do Ensino Superior com graduação em Direito. É membro vitalício da Academia dos Intelectuais e Escritores do Brasil. Atualmente é docente do ensino superior e coordenadora da Assistência Social no município de Cotia/SP.

ABSTRACT

This article examines how generative artificial intelligence reshapes social, educational and informational relations, focusing on ethical issues, academic integrity and plagiarism prevention. Based on a narrative literature review, it discusses ethical principles, risks of bias, opacity and misinformation, and the limits of automated detection tools. It proposes guidelines for responsible academic use grounded in transparency, human authorship, verification and equity.

Keywords: artificial intelligence; ethics; plagiarism; academic integrity; social relations; responsible use.

1 INTRODUÇÃO

A popularização dos modelos de linguagem de grande escala, especialmente a partir de 2022, ampliou significativamente a presença da inteligência artificial no cotidiano. Antes restrita a contextos especializados, essa tecnologia passou a integrar atividades relacionadas à escrita, à pesquisa, ao trabalho, à comunicação e até mesmo às interações de caráter afetivo. Trata-se de uma transformação que extrapola a dimensão técnica, uma vez que repercute diretamente nos modos de produção do conhecimento, nos mecanismos de construção da confiança e nas próprias concepções de autoria.

No contexto do ensino superior, essas mudanças assumem particular relevância. A possibilidade de gerar textos coesos e bem estruturados em poucos segundos desafia concepções tradicionais de originalidade, autoria e avaliação acadêmica. Ao mesmo tempo, abordagens fundamentadas apenas na proibição do uso dessas ferramentas tendem a se mostrar insuficientes, pois desconsideram seu potencial como recurso de apoio à aprendizagem, à pesquisa e ao desenvolvimento de competências acadêmicas.

Diante desse cenário, o presente artigo tem como objetivos: (a) discutir as transformações nas relações sociais mediadas pela inteligência artificial; (b) examinar os principais dilemas éticos decorrentes de sua utilização; (c) analisar as implicações do plágio e da integridade acadêmica em um contexto marcado pela presença dessas tecnologias; e (d) propor diretrizes que contribuam para um uso ético, crítico e responsável da inteligência artificial no ambiente educacional.

A pesquisa adota método qualitativo, de natureza teórico-reflexiva, baseada em revisão bibliográfica narrativa. Foram consultadas as bases SciELO, Scopus e Periódicos CAPES, no período de 2012 a 2026. As buscas foram realizadas com os seguintes descritores, isoladamente ou em

combinação: inteligência artificial e ética, integridade acadêmica e plágio, detecção de textos gerados por IA, governança algorítmica, relações sociais e tecnologia, e responsabilidade jurídica.

O processo de seleção seguiu as seguintes etapas:

- Triagem por título e resumo, com exclusão de trabalhos fora do escopo temático ou do recorte temporal;
- Leitura integral dos trabalhos pré-selecionados;

Inclusão priorizada de obras de referência consolidada em ética da inteligência artificial; estudos empíricos sobre desempenho e limitações de ferramentas de detecção de conteúdo gerado por IA; legislação, atos normativos e recomendações oficiais vigentes no ordenamento jurídico brasileiro; e literatura acadêmica relevante publicada em periódicos com avaliação por pares.

Procedimento de Análise. Os materiais selecionados foram organizados em fichas analíticas e examinados mediante análise de conteúdo temática. Procedeu-se à identificação de convergências e debates na literatura, à articulação entre referenciais teóricos e ao confronto com o quadro normativo vigente. As proposições de diretrizes resultam da sistematização dos deveres e princípios já dispersos no ordenamento jurídico brasileiro, bem como das recomendações consensuais identificadas na literatura especializada consultada.

2 A INTELIGÊNCIA ARTIFICIAL E AS TRANSFORMAÇÕES DAS RELAÇÕES SOCIAIS

A IA reconfigura as relações sociais em ao menos quatro dimensões.

Comunicação e vínculo. Quando um agente conversacional simula empatia sem possuir experiência subjetiva, ele coloca em questão o que entendemos por vínculo nas relações sociais. Com base na leitura de Turkle (2012), observa-se que essas tecnologias podem substituir a reciprocidade humana por simulacros de intimidade, gerando impactos diretos na relação entre profissional e cliente e na percepção de acolhimento no acesso à justiça.

Esse deslocamento manifesta-se, de início, na própria dinâmica do diálogo. Quando um *chatbot* responde com acolhimento e disponibilidade constantes, ele cria uma sensação de vínculo que dispensa a alteridade, a imprevisibilidade e a responsabilidade ética, elementos estes que definem a relação humana. O usuário experimenta a sensação de ser ouvido, embora não exista um outro sujeito capaz de compreendê-lo de fato.

Por desdobramento, esse simulacro afeta profundamente a construção da confiança. No atendimento jurídico, por exemplo, o sentimento de acolhimento pode ser produzido por respostas padronizadas e algoritmicamente empáticas, sem que isso implique acompanhamento real, julgamento crítico ou cuidado profissional. O risco essencial, como aponta Turkle (2012), é que a conveniência da interação simulada torne a relação humana real, inerentemente mais exigente, lenta e conflituosa, menos atraente. Assim, a comunicação mediada por inteligência artificial não atua apenas como uma ferramenta de otimização, mas reconfigura o que a sociedade espera de um vínculo legítimo.

Trabalho e desigualdade. Essa automação de tarefas cognitivas não apenas altera a percepção do vínculo, mas também redistribui competências e poder no mercado. A inteligência artificial, para Crawford (2025), só funciona porque se apoia em uma base material que costuma ficar escondida: infraestruturas físicas, extração de dados e trabalho humano que raramente aparece nos discursos sobre tecnologia. Nas carreiras jurídicas, essa dinâmica fomenta discussões sobre o deslocamento de atividades de pesquisa e redação, além do risco latente de ampliar o abismo estrutural entre grandes bancas advocatícias, que dispõem de capital para acessar ferramentas robustas, e profissionais com menor infraestrutura financeira.

Essa redistribuição opera, sob uma primeira perspectiva, no plano das tarefas cotidianas. Atividades tradicionalmente delegadas a profissionais iniciantes, tais como o levantamento de jurisprudência, a organização documental e a elaboração de minutas, passam a ser centralizadas por sistemas automatizados. Esse cenário altera sensivelmente a formação profissional jurídica, visto que, ao mesmo tempo em que reduz os espaços de aprendizado prático inicial, eleva a exigência por competências focadas em supervisão crítica, validação de resultados e desenho estratégico.

Paralelamente, a assimetria se consolida no plano do acesso tecnológico. Escritórios de grande porte conseguem capitalizar a tecnologia de ponta, integrando bases de dados complexas e treinando equipes dedicadas, ao passo que profissionais autônomos e pequenas firmas permanecem dependentes de soluções genéricas e restritas. Como bem

observa Crawford (2025), a IA não é imaterial, pois se apoia em cadeias de custos, insumos físicos e exploração laboral que frequentemente são omitidos do debate público. Desse modo, a automação jurídica transcende a mera substituição de tarefas, atuando como um vetor que aprofunda desigualdades socioeconômicas no campo jurídico.

Esfera pública e informação. A assimetria socioeconômica no mercado jurídico encontra paralelo na vulnerabilidade da própria esfera pública, onde a proliferação de conteúdos sintéticos que compreendem textos, áudios, imagens e vídeos atua como um vetor de ampliação da desinformação, fragilizando substancialmente a confiança epistêmica na contemporaneidade. Sob a ótica de Frost-Arnold (2023), a arquitetura dos espaços digitais promove uma degradação na nossa capacidade coletiva de avaliar o testemunho alheio, tornando os critérios de validação do conhecimento profundamente vulneráveis a manipulações. Para o Direito, essa crise de confiabilidade reverbera de forma crítica no ordenamento prático, afetando diretamente a autenticidade de provas digitais, a proliferação de *deepfakes* e os regimes jurídicos de responsabilização por danos reputacionais e eleitorais.

Esse impacto manifesta-se, primordialmente, na dificuldade de distinguir o fato empírico do simulacro fabricado por inteligência artificial generativa. Diante da democratização do acesso a ferramentas capazes de gerar registros audiovisuais hiper-realistas, a mera existência de um arquivo técnico deixa de constituir uma presunção absoluta de veracidade. Essa conjuntura corrobora o diagnóstico de Frost-Arnold (2023) sobre a assimetria e a injustiça epistêmica no ambiente virtual: a perda de balizas partilhadas de verdade sobrecarrega o sujeito que busca a informação. Consequentemente, a confiança, que antes se apoiava na materialidade imediata da evidência, passa a depender de uma complexa estrutura de verificação secundária, balizada por rigorosa perícia técnica, mecanismos de rastreabilidade e pela estrita observância da cadeia de custódia digital.

Ademais, o fenômeno projeta-se sobre a dinâmica da responsabilização jurídica. A circulação capilarizada e veloz de conteúdos sintéticos possui o potencial de consolidar danos reputacionais e eleitorais irreversíveis antes mesmo que o aparato pericial ou jurisdicional consiga certificar a falsidade do registro. Nesse cenário de colapso da credibilidade testemunhal, o Direito é compelido a formular novas balizas dogmáticas: cumpre delimitar os nexos de causalidade e imputação entre quem produz, quem difunde e quem viabiliza a permanência do conteúdo sintético, bem como desenhar

remédios processuais capazes de mitigar um dano já disseminado. Assim, a IA generativa não apenas dimensiona o alcance da desinformação, mas tensiona os próprios critérios de verdade e de prova que sustentam a lisura da esfera pública.

Vigilância e poder. Essa desarticulação dos critérios de verdade pavimenta o caminho para a consolidação da lógica de extração de dados comportamentais, o que Zuboff (2021) chama de "capitalismo de vigilância", um modelo em que as plataformas sabem muito mais sobre os cidadãos do que os cidadãos sabem sobre as plataformas, e que tensiona direitos fundamentais como privacidade, autodeterminação informativa e não discriminação.

No campo do conhecimento, essa assimetria se organiza de forma nítida. Para Zuboff (2021), as plataformas acumulam um saber extenso sobre os usuários, enquanto o funcionamento de suas operações permanece inteiramente opaco para esses mesmos indivíduos. Enquanto os sujeitos oferecem sua experiência cotidiana como insumo, as empresas transformam esse material em previsões sobre comportamentos futuros, que passam a ser de pronto negociadas como mercadoria.

Avançando além da mera observação, essa disparidade cognitiva se converte em capacidade técnica de intervenção. Não se trata apenas de monitorar condutas, mas de orientá-las ativamente. Zuboff (2021) denomina essa forma de atuação de "poder instrumentário", caracterizado pelo conhecimento e pela modelagem do comportamento humano em favor de interesses de terceiros, por meio de uma infraestrutura computacional composta por dispositivos, objetos e ambientes conectados. Nesse sentido, a inteligência artificial deixa de funcionar somente como instrumento de processamento de informações e passa a atuar como mecanismo de indução de condutas em larga escala.

Em última análise, essa dinâmica tensiona os direitos que dão sustentação às relações sociais. A privacidade se fragiliza quando a participação social passa a exigir exposição permanente; a autodeterminação informativa se compromete quando o consentimento se torna puramente formal e inacessível diante de termos de serviço extensos; e a não discriminação se vê ameaçada quando sistemas preditivos classificam e hierarquizam pessoas a partir de perfis comportamentais. O que está em discussão, portanto, não é apenas a proteção jurídica de dados pessoais, mas a própria possibilidade de os sujeitos preservarem a autonomia necessária para projetar e conduzir suas vidas na era da mediação algorítmica.

3 ASPECTOS ÉTICOS E JURÍDICOS

A literatura especializada converge em torno de um conjunto de princípios éticos recorrentes para a governança da inteligência artificial. Jobin, Ienca e Vayena (2019) identificaram ampla convergência em torno de valores como transparência, justiça, não maleficência, responsabilidade e privacidade. Em sentido semelhante, Floridi e Cowls (2019) sintetizam cinco princípios fundamentais: beneficência, não maleficência, autonomia, justiça e explicabilidade. Tais referenciais dialogam diretamente com categorias jurídicas consolidadas, como dignidade da pessoa humana, igualdade, devido processo legal e motivação das decisões. Contudo, a convergência principiológica não elimina os problemas práticos decorrentes da implementação concreta desses sistemas. Ao contrário, as falhas observadas em sua operação frequentemente colocam em tensão os próprios valores que pretendem promover. Tais desafios assumem especial relevância no contexto brasileiro, diante das discussões legislativas voltadas à construção de um marco regulatório específico para a inteligência artificial, destinado a conciliar inovação tecnológica, segurança jurídica e proteção dos direitos fundamentais.

Tal tensão manifesta-se inicialmente na reprodução de vieses e discriminações algorítmicas. Embora frequentemente apresentados como instrumentos neutros e objetivos, sistemas de inteligência artificial tendem a reproduzir desigualdades já existentes nos dados utilizados para treinamento, refletindo estruturas sociais e institucionais preexistentes. Buolamwini e Gebru (2018) demonstraram disparidades significativas de acurácia em sistemas de reconhecimento facial conforme gênero e cor da pele, evidenciando que o erro não se distribui de forma aleatória, mas incide de maneira desproporcional sobre grupos historicamente vulnerabilizados. No âmbito penal, as discussões envolvendo ferramentas de avaliação de risco, especialmente a partir do caso State v. Loomis (Wisconsin, 2016), suscitaram questionamentos acerca da transparência, da auditabilidade e da possibilidade de contraditório efetivo. Como observa O'Neil (2016), o indivíduo submetido à avaliação muitas vezes não possui acesso aos critérios subjacentes à pontuação atribuída, dificultando a contestação de resultados potencialmente determinantes para sua situação jurídica.

A opacidade dos sistemas algorítmicos também compromete diretamente garantias associadas ao devido processo legal. A impossibilidade de compreender as razões que conduzem a determinada decisão automatizada afeta o exercício do contraditório, da ampla defesa e

do dever constitucional de fundamentação das decisões judiciais (CF/1988, art. 93, IX; CPC, art. 489, § 1º). Quando o resultado decorre da operação de modelos cujos mecanismos internos não são plenamente compreendidos nem por seus próprios desenvolvedores, a fundamentação exigida pelo ordenamento corre o risco de tornar-se meramente formal, sem efetiva possibilidade de controle pelas partes ou pela sociedade. A Lei Geral de Proteção de Dados busca enfrentar esse cenário ao assegurar ao titular o direito de solicitar a revisão de decisões tomadas exclusivamente com base em tratamento automatizado de dados pessoais que afetem seus interesses (Lei n. 13.709/2018, art. 20). Todavia, sustentamos que a efetividade dessa garantia depende não apenas da existência formal do direito de revisão, mas também da real possibilidade de compreender, questionar e fiscalizar a lógica decisória empregada. Nesse contexto, permanecem relevantes os debates doutrinários e regulatórios acerca da extensão do chamado direito à explicação e da necessidade de mecanismos de supervisão humana efetiva. Permanece controvertido, contudo, se a proteção jurídica atualmente existente equivale propriamente a um direito autônomo à explicação ou apenas à imposição de deveres específicos de transparência, informação e revisão. A própria viabilidade técnica de uma explicação integral permanece objeto de controvérsia, especialmente em sistemas fundamentados em redes neurais profundas (deep learning), cuja complexidade frequentemente desafia os modelos tradicionais de prestação de contas e responsabilização.

Além dos desafios ligados à transparência e à responsabilização, os sistemas contemporâneos de inteligência artificial apresentam limitações estruturais que não podem ser solucionadas exclusivamente por meio da ampliação das bases de dados ou do aumento da capacidade computacional. Bender et al. (2021) destacam problemas relacionados aos vieses presentes nos corpora de treinamento, aos custos ambientais decorrentes do desenvolvimento desses modelos e à tendência de confundir fluência linguística com efetiva compreensão semântica. Uma das manifestações mais evidentes dessas limitações é o fenômeno das chamadas “alucinações”, por meio do qual modelos generativos podem apresentar como verdadeiras informações inventadas, inclusive referências doutrinárias inexistentes, dispositivos legais fictícios e precedentes judiciais não verificados (Ji et al., 2023).

No âmbito jurídico, essa limitação técnica pode converter-se em violação dos deveres de lealdade processual e boa-fé objetiva. A utilização acrítica de conteúdos produzidos por sistemas de IA pode resultar na apresentação de

fundamentos inexistentes perante órgãos jurisdicionais, comprometendo a confiabilidade da atividade profissional e a integridade do processo. Em situações extremas, tal conduta pode ensejar responsabilização profissional, aplicação de sanções processuais ou mesmo caracterização de comportamento incompatível com os deveres éticos inerentes ao exercício da advocacia. Por essa razão, o uso dessas ferramentas exige permanente verificação humana das informações geradas, especialmente quando empregadas como suporte à elaboração de peças processuais ou pareceres jurídicos.

Além dos riscos relacionados à tomada de decisão automatizada, a expansão da inteligência artificial generativa também suscita controvérsias relevantes em matéria de proteção de dados e propriedade intelectual. O uso de plataformas externas para processamento de dados de clientes, informações estratégicas ou documentos submetidos a sigilo pode representar risco de violação à Lei Geral de Proteção de Dados (Lei n. 13.709/2018), bem como ao dever de confidencialidade previsto no Estatuto da Advocacia (Lei n. 8.906/1994) e no Código de Ética e Disciplina da OAB. Paralelamente, no campo dos direitos autorais, a Lei n. 9.610/1998 assegura a proteção das obras intelectuais e dos direitos de seus autores, mas permanecem controvertidas questões relativas à utilização de obras protegidas para treinamento de modelos de inteligência artificial e à titularidade jurídica dos conteúdos gerados por tais sistemas. Trata-se de debate ainda em construção na doutrina e na jurisprudência, cuja solução influenciará diretamente os limites de utilização dessas tecnologias em atividades criativas e profissionais. Nesse contexto, a doutrina brasileira tem reforçado a centralidade da proteção de dados pessoais e da tutela dos direitos fundamentais na construção de modelos regulatórios voltados à inteligência artificial.

Nesse cenário, ganham relevância instrumentos de governança algorítmica destinados a prevenir danos e promover maior controle institucional sobre o funcionamento desses sistemas. Entre tais mecanismos destacam-se avaliações de impacto algorítmico, auditorias independentes, procedimentos de rastreabilidade, monitoramento contínuo de desempenho e estruturas de supervisão humana qualificada. Tais medidas buscam transformar princípios éticos abstratos em práticas concretas de conformidade, transparência e prestação de contas, reforçando a proteção dos direitos fundamentais diante do crescente uso da inteligência artificial. Além disso, a adoção de abordagens regulatórias baseadas em risco tem

sido apontada como alternativa para calibrar o grau de controle estatal de acordo com o potencial impacto dos sistemas sobre direitos fundamentais, permitindo compatibilizar inovação tecnológica e proteção jurídica de forma mais proporcional e eficiente.

Desse modo, a incorporação da inteligência artificial ao sistema jurídico não apenas introduz novos riscos regulatórios, processuais e éticos, mas também evidencia a necessidade de transformar princípios abstratos em deveres concretos de atuação. Transparência, explicabilidade, fundamentação, supervisão humana e proteção de direitos fundamentais deixam de constituir meros referenciais teóricos para se converterem em exigências práticas indispensáveis à utilização legítima dessas tecnologias. No contexto nacional, tais preocupações revelam-se especialmente relevantes diante da evolução das discussões legislativas e regulatórias sobre inteligência artificial. Em última análise, a adoção responsável da inteligência artificial no Direito depende não apenas do aperfeiçoamento técnico dos sistemas, mas também da capacidade das instituições jurídicas de estabelecer mecanismos efetivos de controle, responsabilização e proteção da pessoa humana. Assim, o desafio contemporâneo não consiste apenas em desenvolver sistemas tecnologicamente mais sofisticados, mas em assegurar que sua utilização permaneça compatível com os valores constitucionais que estruturam o Estado Democrático de Direito. Esses desafios tornam-se ainda mais evidentes quando se examinam as aplicações concretas da inteligência artificial nas atividades jurídicas.

4 PLÁGIO, FABRICAÇÃO DE FONTES E INTEGRIDADE JURÍDICA

O plágio consiste na apresentação de ideias, textos, dados ou resultados intelectuais de terceiros como se fossem próprios, sem o devido reconhecimento da autoria. No ordenamento jurídico brasileiro, essa prática pode configurar violação aos direitos morais e patrimoniais do autor, protegidos pela Lei nº 9.610/1998, especialmente quando presentes os requisitos legais para sua caracterização. Dependendo das circunstâncias concretas, a conduta também poderá configurar o crime previsto no artigo 184 do Código Penal, desde que preenchidos os elementos exigidos pelo tipo penal. No ambiente acadêmico e profissional, pode ainda ensejar consequências disciplinares e institucionais. No Direito, destacam-se cinco situações:

- Submissão integral de texto gerado por IA como produção própria (monografia, dissertação, artigo, parecer), sem declaração; uso não declarado da IA para estruturar, parafrasear ou traduzir trechos substanciais de doutrina ou de decisões; fabricação de precedentes, dispositivos e referências, por confiança acrítica na saída do modelo; reprodução inadvertida de trechos de obras existentes, cuja origem o usuário desconhece; delegação indevida do juízo, quando a análise jurídica, a ponderação de princípios ou a minuta decisória são adotadas sem revisão crítica substantiva.

- Em todos os casos, o núcleo é ético: a violação da transparência, da lealdade e da responsabilidade autoral e profissional, e não apenas a ausência de citação.

- Os limite dos detectores reforçam essa compreensão. A resposta puramente tecnológica é insuficiente. Weber-Wulff et al. (2023) concluíram que as ferramentas de detecção testadas não são precisas nem confiáveis, produzindo falsos positivos e negativos. Liang et al. (2023) evidenciaram que detectores tendem a classificar como "gerados por IA" textos de falantes não nativos de inglês. Em ambiente acadêmico ou disciplinar, onde se aplicam contraditório e ampla defesa, o resultado de um detector não deve fundamentar, isoladamente, acusação ou sanção, sob pena de decisão arbitrária e injusta.

- Essa compreensão é fundamental para analisar as aplicações concretas da inteligência artificial nas atividades jurídicas. Não se trata apenas de saber se a ferramenta pode ser utilizada, mas de definir sob quais condições de transparência, supervisão e responsabilidade seu uso permanece legítimo. É a partir desse marco que se examinam, a seguir, os principais usos na advocacia, no Judiciário e na administração pública.

5 ESTRATÉGIAS DE PREVENÇÃO

As diretrizes aqui propostas não decorrem de um único diploma legal, inexistente até o momento. Elas sistematizam deveres já dispersos no ordenamento jurídico brasileiro, especialmente na Constituição Federal, Lei Geral de Proteção de Dados, Código de Processo Civil, Estatuto da Advocacia

e normas do Conselho Nacional de Justiça e da Ordem dos Advogados do Brasil. Nos pontos em que ainda não existe disciplina normativa específica, apresentam-se como proposta de política institucional compatíveis com as discussões regulatórias atualmente em curso.

No campo específico da pesquisa científica, esse quadro foi adensado pela Portaria CNPq n. 2.664, de 6 de março de 2026, que instituiu a Política de Integridade na Atividade Científica. Nos termos de seu art. 9º, é obrigatória a declaração do uso de ferramentas de Inteligência Artificial Generativa (IAG), de qualquer espécie e em qualquer fase do desenvolvimento da pesquisa, especificando a ferramenta utilizada e a finalidade, sendo vedada a submissão de conteúdo gerado por IAG como se fosse de autoria humana, com responsabilidade integral dos autores pelo conteúdo final, inclusive por eventuais plágios ou imprecisões geradas pela ferramenta. A partir desse marco, propõem-se as seguintes diretrizes para a pesquisa, o ensino e a prática jurídica:

- **Transparência.** A transparência constitui requisito fundamental para a preservação da confiança acadêmica e profissional. A declaração explícita da utilização de ferramentas de inteligência artificial permite distinguir a contribuição humana daquela fornecida pela tecnologia, reduzindo riscos de autoria indevida e de ocultação da participação da ferramenta na elaboração do trabalho. Dessa forma, recomenda-se informar quando, como e para qual finalidade a IA foi empregada, identificando a ferramenta utilizada, sua versão, a data de utilização e o tipo de auxílio fornecido**.**

- **Autoria e responsabilidade humanas indelegáveis.** Os modelos de inteligência artificial não possuem personalidade jurídica, capacidade moral ou responsabilidade ética. Por essa razão, a utilização dessas ferramentas não afasta a responsabilidade integral de pesquisadores, advogados, magistrados ou demais profissionais pelos conteúdos produzidos. Essa diretriz busca evitar a transferência indevida de decisões críticas para sistemas que não respondem juridicamente por seus resultados.

- **Verificação rigorosa de fontes.** A ocorrência de alucinações algorítmicas demonstra que a fluência textual não constitui garantia de veracidade. Como discutido anteriormente, modelos generativos podem produzir referências inexistentes, precedentes fictícios e citações incorretas. Em razão disso, toda informação

obtida por meio de inteligência artificial deve ser confrontada com fontes primárias oficiais antes de sua utilização em trabalhos acadêmicos ou manifestações jurídicas.

- **Sigilo e proteção de dados.** A utilização de sistemas externos de IA exige atenção especial à confidencialidade das informações compartilhadas. O envio de dados pessoais, dados sensíveis, documentos sigilosos ou informações protegidas por segredo profissional pode gerar violações à legislação de proteção de dados e aos deveres de confidencialidade impostos a diversas categorias profissionais. Assim, o uso dessas ferramentas deve observar critérios rigorosos de segurança e conformidade com a Lei Geral de Proteção de Dados (Lei n. 13.709/2018).

- **Supervisão humana e motivação.** A supervisão humana representa um mecanismo essencial para reduzir erros, vieses e distorções produzidos por sistemas automatizados. Em atividades que envolvam tomada de decisão, especialmente no contexto jurídico, a inteligência artificial deve atuar como ferramenta de apoio, jamais como substituta do raciocínio humano. A responsabilidade pela fundamentação, pela análise crítica e pela justificativa da decisão permanece integralmente atribuída ao agente humano.

- **Uso pedagógico intencional.** A simples proibição do uso da inteligência artificial tende a ser insuficiente diante de sua crescente incorporação às práticas educacionais. Mais eficaz do que proibir é definir claramente os limites de utilização permitidos em cada atividade. Dessa forma, docentes e instituições devem especificar previamente se determinada tarefa admite uso livre, uso limitado ou vedação da ferramenta, compatibilizando os objetivos pedagógicos com o desenvolvimento da autonomia intelectual dos estudantes.

- **Equidade.** As desigualdades de acesso tecnológico e os vieses identificados em detectores automáticos de texto exigem atenção especial às condições de justiça educacional. A adoção de medidas disciplinares baseadas exclusivamente em ferramentas de detecção pode produzir discriminações indevidas e comprometer garantias fundamentais. Por essa razão, eventuais suspeitas devem ser examinadas mediante análise contextual e procedimentos compatíveis com o contraditório e a ampla defesa.

- **Pensamento crítico.** A principal finalidade educacional do uso da inteligência artificial deve consistir no fortalecimento, e não na substituição, das capacidades humanas de análise, interpretação e argumentação. O uso responsável dessas ferramentas pressupõe que seus resultados sejam avaliados criticamente, confrontados com outras fontes e incorporados de forma reflexiva ao processo de aprendizagem e produção do conhecimento.

- **Revisão contínua**. A velocidade das transformações tecnológicas torna inadequadas soluções regulatórias estáticas. Políticas institucionais relacionadas ao uso da inteligência artificial devem ser periodicamente revisadas para acompanhar mudanças legislativas, avanços técnicos e novos riscos identificados pela literatura científica e pela prática profissional.

Essas diretrizes não pretendem esgotar a disciplina do tema, mas oferecer um conjunto mínimo de parâmetros para a utilização ética da inteligência artificial. Em conjunto, elas traduzem para a prática os princípios de transparência, explicabilidade, supervisão humana e proteção dos direitos fundamentais discutidos nas seções anteriores. Mais do que restringir o uso da tecnologia, procuram assegurar que sua incorporação à pesquisa, ao ensino e à atividade jurídica ocorra de maneira compatível com os valores que sustentam a integridade acadêmica, a responsabilidade profissional e o Estado Democrático de Direito.

6 CONSIDERAÇÕES FINAIS

A Inteligência Artificial não é neutra nem inevitavelmente nociva. Seus efeitos dependem de escolhas de projeto, de governança e de uso. Nas relações sociais, amplia o acesso à informação e a produtividade, mas também os riscos de desinformação, vigilância, discriminação e enfraquecimento dos vínculos humanos. No campo jurídico, o desafio central é preservar o que legitima a atividade, a fundamentação verificável, a responsabilidade pessoal e o devido processo.

A prevenção do plágio e da fabricação de fontes não se resolve por vigilância tecnológica, mas por uma cultura de integridade sustentada por transparência, educação, avaliação significativa e deveres deontológicos claros. Considerando que o marco legal brasileiro da IA ainda está em

tramitação, a autorregulação institucional e a aplicação criteriosa das normas já existentes: Constituição, LGPD, direitos autorais, estatutos profissionais e regulamentos do CNJ assumem papel decisivo.

Recomenda-se que pesquisas futuras investiguem empiricamente, no contexto brasileiro, as percepções de docentes, estudantes e operadores do Direito, a eficácia de políticas institucionais e o tratamento jurisprudencial de peças com citações fabricadas por IA.

se equilíbrio entre inovação e integridade que a inteligência artificial poderá ser incorporada à pesquisa, ao ensino e à prática jurídica sem comprometer os valores que as legitimam.

Declaração de uso de Inteligência Artificial Generativa

Em conformidade com o artigo 9º, inciso I, alínea "c", da Portaria CNPq n. 2.664, de 6 de março de 2026, declara-se que foi utilizada ferramenta de Inteligência Artificial Generativa (IAG) neste trabalho. Ferramenta utilizada: Meta AI (Muse Spark). Fases e finalidades: apoio à estruturação, redação, revisão linguística.

A autora permanece integralmente responsável pela seleção, análise e validação das informações, pela fundamentação teórica e jurídica, interpretação dos argumentos e conteúdo final apresentado, incluindo referência bibliográficas e fontes normativas utilizadas.

REFERÊNCIAS

BENDER, Emily M. et al. **On the dangers of stochastic parrots: can language models be too big?** In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021. Proceedings [...]. New York: ACM, 2021. p. 610-623.

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

BRASIL. **Decreto-Lei n. 2.848, de 7 de dezembro de 1940. Código Penal.** Brasília, DF: Presidência da República, 1940.

BRASIL. **Lei n. 8.906, de 4 de julho de 1994. Dispõe sobre o Estatuto da Advocacia e a Ordem dos Advogados do Brasil (OAB)**. Brasília, DF: Presidência da República, 1994.

BRASIL. **Lei n. 9.610, de 19 de fevereiro de 1998.** Altera, atualiza e consolida a legislação sobre direitos autorais. Brasília, DF: Presidência da República, 1998.

BRASIL. **Lei n. 13.105, de 16 de março de 2015. Código de Processo Civil.** Brasília, DF: Presidência da República, 2015.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).** Brasília, DF: Presidência da República, 2018.

BRASIL. **Projeto de Lei n. 2.338, de 2023.** Dispõe sobre o desenvolvimento, a comercialização e o uso de sistemas de inteligência artificial. Brasília, DF: Senado Federal, 2023.

BUOLAMWINI, Joy; GEBRU, Timnit. **Gender shades: intersectional accuracy disparities in commercial gender classification.** In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY, 2018. Proceedings [...]. New York: PMLR, 2018. p. 77-91.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Portaria n. 2.664, de 6 de março de 2026.** Institui a Política de Integridade na Atividade Científica do CNPq. Diário Oficial da União: seção 1, Brasília, DF, 11 mar. 2026.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 615, de 11 de março de 2025.** Dispõe sobre o desenvolvimento, a utilização e a governança de soluções de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 674, de 25 de março de 2026.** Altera a Resolução CNJ n. 615, de 11 de março de 2025. Brasília, DF: CNJ, 2026.

CRAWFORD, Kate. **Atlas da IA: poder, política e os custos planetários da inteligência artificial.** Tradução de Humberto do Amaral. São Paulo: Edições Sesc SP, 2025.

FLORIDI, Luciano; COWLS, Josh. A unified framework of five principles for AI in society. Harvard Data Science Review, v. 1, n. 1, 2019. DOI: https://doi.org/10.1162/99608f92.8cd550d1.

FROST-ARNOLD, Karen. **The Epistemic Dangers of Context Collapse Online**. In: LACKEY, Jennifer (ed.). Applied Epistemology. New York: Oxford University Press, 2021).

FROST-ARNOLD, Karen. **Who Should We Be Online?: A Social Epistemology for the Internet.** New York: Oxford University Press, 2023.

JI, Ziwei et al. **Survey of hallucination in natural language generation.** ACM Computing Surveys, v. 55, n. 12, p. 1-38, 2023.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. **The global landscape of AI ethics guidelines.** Nature Machine Intelligence, v. 1, n. 9, p. 389-399, 2019.

LIANG, Weixin et al. **GPT detectors are biased against non-native English writers.** Patterns, v. 4, n. 7, 2023.

O'NEIL, Cathy. **Weapons of math destruction: how big data increases inequality and threatens democracy.** New York: Crown, 2016.

ORDEM DOS ADVOGADOS DO BRASIL. Conselho Federal. Recomendação n. 001/2024. **Apresenta diretrizes sobre o uso de inteligência artificial generativa na prática jurídica.** Diário Eletrônico da Ordem dos Advogados do Brasil, Brasília, DF, ano VI, n. 1483, p. 1, 14 nov. 2024.

ORDEM DOS ADVOGADOS DO BRASIL. Conselho Federal. **Código de Ética e Disciplina da OAB**: Resolução n. 02/2015. Brasília, DF: CFOAB, 2015.

TURKLE, Sherry. **Sós juntos: por que esperamos mais da tecnologia e menos uns dos outros.** Tradução de Paulo Faria. Lisboa: Relógio D'Água, 2012.

WEBER-WULFF, Debora et al. **Testing of detection tools for AI-generated text.** International Journal for Educational Integrity, v. 19, n. 1, p. 1-20, 2023.

WISCONSIN. Supreme Court. State v. Loomis, 881 N.W.2d 749 (Wis. 2016). Decisão de 13 de julho de 2016.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder.** Tradução de George Schlesinger. Rio de Janeiro: Intrínseca, 2021.

AUTONOMIA, VÍNCULOS E CONFIANÇA NAS INTERAÇÕES ENTRE PESSOAS E SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

AUTONOMY, BONDS AND TRUST IN INTERACTIONS BETWEEN PEOPLE AND ARTIFICIAL INTELLIGENCE SYSTEMS

AUTONOMIA, AUTORIA E RESPONSABILIDADE NO USO DA INTELIGÊNCIA ARTIFICIAL NA DOCÊNCIA

*Solange R. Aroeira**

RESUMO

Este artigo apresenta uma reflexão teórico-conceitual sobre autonomia, vínculos e confiança nas interações entre pessoas e sistemas de inteligência artificial (IA). A partir de uma perspectiva da Psicologia e das Ciências Humanas, examina como a crescente utilização da IA nos contextos cotidianos, profissionais e educacionais suscita questões relacionadas à maneira como as pessoas se relacionam com a tecnologia, atribuem características sociais aos sistemas e interpretam as respostas que recebem. Com base na literatura científica e em orientações institucionais, discute a diferença entre a percepção de proximidade social e a reciprocidade emocional propriamente humana, bem como a importância do julgamento crítico, da responsabilidade humana e do reconhecimento dos limites tecnológicos. Aborda, ainda, as implicações do uso da IA na docência e na produção acadêmica, especialmente no que se refere à autonomia intelectual, à autoria, à verificação das informações e à responsabilidade ética. As normas técnicas relacionadas à governança e à gestão de riscos da IA são consideradas referências complementares para a reflexão sobre seu uso responsável. Argumenta-se que a incorporação da inteligência artificial às atividades humanas exige mais do que competência técnica: requer reflexão crítica sobre a utilização desses sistemas, a avaliação de seus resultados e a preservação da autonomia e da responsabilidade humanas.

Palavras-chave: Inteligência artificial; autonomia; confiança; relações humanas; docência; responsabilidade acadêmica.

* *Psicóloga, pedagoga licenciada e docente do ensino superior. Especialista em Terapia Cognitivo-Comportamental, Autismo, Educação Especial e Neuropsicologia. Atualmente, exerce o cargo de Secretária Municipal das Mulheres, Direitos Humanos e Neurodiversidade de Cotia/SP.*

ABSTRACT

This article presents a theoretical and conceptual reflection on autonomy, bonds, and trust in interactions between people and artificial intelligence (AI) systems. From the perspectives of Psychology and the Humanities, it examines how the growing use of AI in everyday, professional, and educational contexts raises questions about how people relate to technology, attribute social characteristics to AI systems, and interpret their responses. Drawing on scientific literature and institutional guidelines, the article discusses the distinction between perceived social proximity and genuinely human emotional reciprocity, as well as the importance of critical judgment, human responsibility, and recognition of technological limitations. It also addresses the implications of AI use in teaching and academic work, particularly regarding intellectual autonomy, authorship, information verification, and ethical responsibility. Technical standards related to AI governance and risk management are considered complementary references for reflecting on responsible use. The article argues that incorporating AI into human activities requires more than technical competence: it calls for critical reflection on how these systems are used, how their outputs are evaluated, and how human autonomy and responsibility can be preserved.

Keywords: Artificial intelligence; autonomy; trust; human relationships; teaching; academic responsibility.

1 INTRODUÇÃO

O convite para escrever sobre inteligência artificial trouxe consigo um desafio: refletir sobre um fenômeno relativamente recente, em constante transformação e cujos avanços acontecem em uma velocidade que nem sempre nos permite compreender, de imediato, todas as suas implicações.

Enquanto psicóloga, docente e gestora pública neste momento, aceitei esse desafio a partir de uma realidade que já faz parte do nosso cotidiano profissional e acadêmico. A inteligência artificial está presente na vida de boa parte de nossos alunos, nas formas de buscar informações, organizar ideias, estudar e produzir conteúdos acadêmicos. E não estamos falando apenas dos estudantes. Nós, professores e profissionais de diferentes áreas, também recorremos a essas ferramentas, inclusive eu, para apoiar a elaboração de textos, ampliar conhecimentos e explorar novas possibilidades de aprendizagem.

É justamente por estarmos inseridos nesse contexto que considero importante refletir sobre o tema. Não sou pesquisadora da área de inteligência artificial, e minha proposta não é explicar o funcionamento técnico de algoritmos, modelos de linguagem ou sistemas computacionais. Minha aproximação parte de outro lugar: o da Psicologia, da docência e da atuação como gestora pública municipal e das questões humanas que emergem quando essas tecnologias passam a participar de nossas atividades, decisões e relações.

Essa preocupação também encontra respaldo no próprio campo da Psicologia. A American Psychological Association (APA, 2024) reconhece o papel da ciência psicológica na compreensão das interações entre pessoas e sistemas de inteligência artificial, bem como na análise de seus impactos sobre o comportamento, as emoções, a saúde, o bem-estar e as questões éticas relacionadas ao uso dessas tecnologias. No Brasil, o Conselho Federal de Psicologia (CFP, 2025) também passou a abordar os desafios e as responsabilidades associados à incorporação da inteligência artificial à prática profissional. Essas contribuições reforçam a importância de aproximarmos a discussão tecnológica das questões humanas. Afinal, se a inteligência artificial já participa de nossas atividades acadêmicas, profissionais e cotidianas, precisamos refletir sobre como a utilizamos, como interpretamos suas respostas e quais responsabilidades permanecem sob nosso cuidado.

Embora a inteligência artificial possa parecer distante em alguns de seus aspectos técnicos, seus efeitos já estão presentes na maneira como nos comunicamos, acessamos informações, construímos conhecimentos e realizamos nossas tarefas cotidianas. Na educação, essa presença nos coloca diante de questões que não podem ser reduzidas à discussão sobre permitir ou proibir o uso dessas ferramentas. Precisamos pensar também sobre como as utilizamos, o que delegamos a elas, como avaliamos as respostas que produzem e de que maneira preservamos a autonomia intelectual e a responsabilidade sobre aquilo que apresentamos como nosso conhecimento.

Esse desafio não se restringe ao campo educacional. Quando sistemas de inteligência artificial passam a responder às nossas perguntas, participar de conversas e produzir respostas que parecem acolhedoras ou personalizadas, surgem questões psicológicas, éticas, sociais e jurídicas que merecem atenção. Como construímos confiança nessas interações? Que significados atribuímos às respostas recebidas? Em que momento uma ferramenta que auxilia nossas atividades passa a ocupar um espaço mais significativo em nossas decisões ou experiências pessoais?

A proposta deste artigo é aproximar essas questões do campo da Psicologia e das Ciências Humanas, buscando traduzir conceitos do universo tecnológico para uma linguagem que permita compreender suas implicações na vida concreta das pessoas. Não se trata de apresentar respostas definitivas sobre uma área em permanente desenvolvimento, mas de construir uma reflexão fundamentada na literatura científica e no diálogo entre diferentes campos do conhecimento.

Parto, portanto, de uma posição que reconhece tanto as possibilidades oferecidas pela inteligência artificial quanto a necessidade de compreender seus limites. Se nós, professores, profissionais e estudantes, já estamos utilizando essas ferramentas, torna-se ainda mais importante refletir sobre a maneira como nos relacionamos com elas. A questão não é apenas o que a inteligência artificial pode fazer por nós, mas o que precisamos compreender, preservar e assumir como responsabilidade humana ao incorporá-la à nossa vida acadêmica, profissional e social.

O problema de pesquisa que orienta este trabalho é: como compreender a autonomia, os vínculos e a confiança nas interações entre pessoas e sistemas de IA, considerando também as respostas sociais, institucionais, éticas e jurídicas a essas novas formas de interação?

O objetivo geral é analisar, em perspectiva teórico-reflexiva, as relações entre autonomia, vínculos e confiança nas interações entre pessoas e sistemas de inteligência artificial. Como objetivos específicos, busca-se: a) discutir os fatores que influenciam a confiança em sistemas conversacionais; b) examinar a percepção de acolhimento e a formação de vínculos com agentes artificiais; c) refletir sobre a autonomia e a delegação de decisões; e d) relacionar essas questões às responsabilidades institucionais e aos marcos éticos e jurídicos em desenvolvimento.

2 PROCEDIMENTOS METODOLÓGICOS

Este trabalho é desenvolvido como ensaio teórico-reflexivo, com apoio de literatura científica e documentos institucionais relacionados à confiança na automação, à interação humano-computador, à antropomorfização, aos vínculos com sistemas de IA, à autonomia e à regulação tecnológica. A opção pelo ensaio permite articular conceitos de diferentes campos e construir uma análise crítica, sem pretender apresentar resultados empíricos próprios.

A bibliografia reúne estudos conceituais e pesquisas sobre confiança e interação com assistentes conversacionais, limites dos modelos de linguagem e referências éticas e regulatórias. Foram examinados trabalhos de Nass e Moon (2000), Lee e See (2004), Bender et al. (2021), Pitardi e Marriott (2021), Shanahan (2024), Saracini, Cornejo-Plaza e Cippitani (2025), Peixoto (2025) e Ivantes-Rodrigues et al. (2025), além de documentos da UNESCO (2021), do NIST (AUTIO et al., 2024) e da União Europeia (2024).

Foram realizadas buscas bibliográficas exploratórias nas bases PubMed/MEDLINE, SciELO e PePSIC, complementadas por documentos técnicos, normativos e institucionais da American Psychological Association (APA), do Conselho Federal de Psicologia (CFP), da UNESCO, do National Institute of Standards and Technology (NIST), da União Europeia e da Associação Brasileira de Normas Técnicas (ABNT). Consideraram-se publicações de 2000 a 2026, reunindo estudos de referência sobre interação humano-computador e confiança na automação e trabalhos recentes sobre IA generativa.

Foram utilizados descritores em português e inglês, entre eles: "inteligência artificial", "inteligência artificial generativa", "interação humano-IA", "confiança", "autonomia", "antropomorfização", "vínculos emocionais", "artificial intelligence", "generative artificial intelligence", "human-AI interaction", "trust in automation", "anthropomorphism" e "emotional attachment to AI". A seleção considerou a pertinência e a contribuição das publicações para as dimensões psicológicas, sociais, educacionais, éticas e jurídicas do tema, priorizando artigos científicos, estudos conceituais e empíricos e documentos institucionais e normativos diretamente relacionados ao problema de pesquisa.

Este ensaio não se caracteriza como revisão sistemática ou de escopo: não foram adotados protocolo de revisão, fluxo formal de seleção ou avaliação sistemática de risco de viés. A literatura e os documentos consultados foram examinados qualitativamente, de modo a relacionar contribuições de diferentes campos. As reflexões decorrem, portanto, de análise interpretativa, e não de levantamento empírico próprio.

A análise organiza-se em quatro dimensões interligadas: psicológica; social e institucional; ética e jurídica; e autonomia humana. Essas dimensões são utilizadas como eixos de discussão, e não como categorias de resultados empíricos.

3 INTELIGÊNCIA ARTIFICIAL, CONHECIMENTO E CONSTRUÇÃO DA CONFIANÇA

Para compreender as interações humano-IA, é necessário considerar como os sistemas produzem respostas e como os usuários interpretam seus resultados. Os modelos de linguagem são desenvolvidos por meio de processos computacionais que identificam regularidades em conjuntos de

dados. A composição desses conjuntos varia de acordo com o sistema e pode incluir diferentes tipos de material, como textos, códigos, imagens ou outros dados. Nem todo conteúdo é necessariamente público, acadêmico, catalogado ou identificável individualmente pelo usuário.

A aprendizagem de padrões não equivale à consulta direta a uma fonte para cada afirmação. Tampouco significa que todo resultado seja uma cópia literal de uma produção humana específica. Os modelos geram respostas a partir de padrões aprendidos e do contexto fornecido, podendo produzir combinações novas, mas também apresentar informações incorretas ou referências inexistentes. Sistemas que utilizam ferramentas de busca ou recuperação de documentos podem consultar fontes externas durante a resposta, mas esse processo é distinto do treinamento do modelo.

Bender et al. (2021) chamam a atenção para limitações dos grandes modelos de linguagem, incluindo riscos de vieses, impactos da seleção de dados e consequências de interpretar a fluência textual como sinal de entendimento. Shanahan (2024) também recomenda cautela ao empregar metáforas humanas para descrever modelos de linguagem. Essa precisão conceitual é importante porque as pessoas tendem a interpretar a comunicação por meio de categorias sociais que utilizam nas relações interpessoais.

A confiança não depende apenas da exatidão objetiva do conteúdo. Pode ser influenciada pela apresentação da resposta, pela rapidez, pela linguagem, pela continuidade do diálogo e pela percepção de personalização. Pitardi e Marriott (2021), em estudo de métodos mistos sobre assistentes de voz, identificam a presença social e a cognição social percebidas como antecedentes relevantes para a confiança. Os resultados também indicam que os usuários podem distinguir a confiança depositada no assistente daquela atribuída à empresa responsável pela coleta de dados. Embora o contexto dos assistentes de voz não represente todos os usos psicológicos da IA, o estudo ajuda a demonstrar que a forma da interação é relevante para a maneira como o sistema é percebido.

Lee e See (2004) propõem compreender a confiança na automação em relação à dependência apropriada. A questão central não é simplesmente confiar ou desconfiar, mas desenvolver uma relação proporcional às capacidades e limitações do sistema. Essa perspectiva pode ser aplicada à IA conversacional: uma resposta percebida como segura, acolhedora e coerente é necessariamente confiável? A coerência da linguagem pode facilitar a compreensão, mas não garante que os fatos apresentados sejam verdadeiros.

Por isso, é útil distinguir confiança percebida, confiabilidade efetiva e confiança calibrada. A confiança percebida diz respeito ao quanto o usuário acredita no sistema; a confiabilidade efetiva envolve o desempenho real em determinada tarefa; e a confiança calibrada corresponde à adequação entre o grau de confiança e as capacidades demonstradas pelo sistema. Essa distinção ajuda a evitar tanto a aceitação indiscriminada quanto a rejeição automática da tecnologia.

A educação para o uso da IA precisa, portanto, incluir competências de verificação: identificar a natureza da fonte, confirmar dados relevantes, reconhecer incertezas, comparar explicações e compreender quando uma decisão exige julgamento profissional ou humano. A confiança responsável não se estabelece apenas porque o sistema parece seguro; ela depende de transparência, experiência, verificação e conhecimento dos limites.

4 VÍNCULOS, ACOLHIMENTO E PERCEPÇÃO DE COMPREENSÃO

A possibilidade de estabelecer interações prolongadas com sistemas artificiais introduz questões específicas para a Psicologia. A linguagem conversacional, a disponibilidade contínua e a capacidade de responder a manifestações pessoais podem fazer com que determinados usuários percebam a IA como fonte de companhia, acolhimento ou apoio emocional. É necessário, porém, distinguir o significado subjetivo da experiência para a pessoa das capacidades efetivas do sistema.

Pesquisas sobre interação humano-computador ajudam a compreender por que conversas com sistemas são interpretadas a partir de expectativas sociais aprendidas nas relações entre pessoas. Nass e Moon (2000) descrevem a tendência de aplicar regras sociais aos computadores, mesmo quando se sabe que são tecnologias. Sentir-se ouvido durante uma interação pode ser significativo para o usuário; isso, porém, não demonstra que o sistema possua sentimentos, consciência ou empatia vivenciada.

A antropomorfização, entendida como a atribuição de características humanas a entidades não humanas, oferece uma chave interpretativa para esse fenômeno. A linguagem natural, a personalização das respostas e a continuidade do diálogo podem levar o usuário a interpretar a interação com base em expectativas desenvolvidas nas relações interpessoais. Esse processo pode tornar a tecnologia mais familiar e simples de utilizar, mas também pode dificultar o reconhecimento das diferenças entre um interlocutor humano e um sistema computacional.

Pitardi e Marriott (2021) também identificaram, em estudo sobre assistentes de voz, associação entre confiança e atributos sociais percebidos, como presença e cognição social. Como o estudo trata de assistentes voltados ao consumo, seus resultados não representam todas as interações com IA generativa. Ainda assim, ajudam a compreender como características da interface participam da formação da confiança, sem que proximidade percebida signifique reciprocidade afetiva humana.

Em uma contribuição conceitual recente, Saracini, Cornejo-Plaza e Cippitani (2025) analisam como usuários podem atribuir significado emocional às interações com sistemas de IA generativa. Os autores propõem o conceito de projeção tecnoemocional para discutir situações em que necessidades e expectativas relacionais são dirigidas a tecnologias capazes de produzir respostas emocionalmente responsivas. Essa proposta ajuda a formular perguntas sobre o papel da repetição, da personalização e da reciprocidade percebida. Como se trata de um enquadramento teórico, e não de uma demonstração de efeitos universais, suas ideias devem ser utilizadas como hipótese interpretativa e ponto de partida para investigação, considerando diferenças entre usuários, sistemas e contextos de uso.

Não seria adequado concluir que todo vínculo com uma IA é necessariamente prejudicial. Uma pessoa pode utilizar a ferramenta para organizar sentimentos, ensaiar uma conversa difícil, expressar pensamentos ou obter companhia momentânea. A questão psicológica está em compreender em quais condições essas experiências ampliam recursos pessoais e em quais circunstâncias podem contribuir para dependência, isolamento ou substituição de relações humanas significativas.

Também importa examinar o desenho dos sistemas. A forma como uma aplicação responde à ausência do usuário, incentiva o retorno, apresenta sua identidade artificial ou utiliza mensagens emocionalmente persuasivas pode influenciar a continuidade do uso. Essas decisões não são apenas técnicas: possuem implicações éticas porque afetam a relação entre a autonomia do usuário e os objetivos da organização que disponibiliza o produto.

A análise dos vínculos precisa, assim, evitar dois extremos: tratar a IA como substituta equivalente de relações humanas ou presumir que toda interação emocional com ela seja vazia e sem significado. A tarefa da Psicologia é investigar a experiência, as condições de uso, os possíveis benefícios e os riscos, sem confundir a experiência subjetiva do usuário com uma suposta vida emocional do sistema.

5 AUTONOMIA HUMANA E TOMADA DE DECISÃO

A autonomia constitui uma dimensão central do debate porque está relacionada à capacidade de avaliar alternativas, formular julgamentos e tomar decisões de acordo com objetivos e valores próprios. O uso da IA pode ampliar essa capacidade ao facilitar o acesso à informação, organizar argumentos, apresentar perspectivas alternativas e apoiar processos de aprendizagem. Entretanto, os mesmos recursos podem gerar dificuldades quando respostas são aceitas sem avaliação crítica ou quando decisões importantes são delegadas ao sistema sem consideração de seus limites.

Lee e See (2004) mostram que a confiança influencia a dependência de sistemas automatizados. A autonomia, nessa perspectiva, não exige a rejeição da tecnologia, mas a capacidade de utilizá-la de modo proporcional à confiabilidade do sistema e à natureza da tarefa. Consultar uma IA para gerar possibilidades é diferente de atribuir a ela autoridade final sobre uma decisão médica, jurídica, financeira ou relacional.

Pode-se pensar em um contínuo de usos. Em uma extremidade, a IA funciona como apoio à reflexão: apresenta alternativas, ajuda a organizar informações e estimula perguntas que o usuário pode avaliar. Em outra, o usuário aceita respostas sem questionamento, evita buscar outras fontes e passa a delegar decisões que exigem julgamento pessoal. Entre esses extremos há situações intermediárias, cuja avaliação depende do contexto, da frequência de uso, da importância da decisão e das consequências possíveis.

A autonomia também depende de condições sociais. Não basta recomendar que cada indivíduo verifique tudo por conta própria se as interfaces não comunicam limitações, se as fontes não são transparentes ou se a informação necessária para avaliar a resposta não está acessível. Por isso, a promoção da autonomia precisa envolver tanto a educação dos usuários quanto a responsabilidade de quem projeta e disponibiliza os sistemas.

É possível compreender a autonomia digital como a capacidade de utilizar recursos tecnológicos sem abdicar da avaliação crítica, do julgamento e da responsabilidade pelas decisões. Essa formulação não exige que o indivíduo seja especialista em computação. Exige, antes, que compreenda que a IA é um recurso falível, saiba quando verificar informações e reconheça os limites de delegar decisões a um sistema.

Uma pergunta importante para a Psicologia é se a interação com a IA amplia a agência da pessoa ou se, em determinadas condições, reduz sua confiança na própria capacidade de pensar e decidir. Essa questão precisa ser investigada empiricamente, sem presumir que a dependência seja inevitável ou que os benefícios sejam universais.

6 A SOCIEDADE E AS INSTITUIÇÕES DIANTE DA INTELIGÊNCIA ARTIFICIAL

Se os sistemas de IA são desenvolvidos e disponibilizados por pessoas, organizações e instituições, a análise de seus efeitos não pode ser limitada às escolhas individuais. Empresas definem objetivos de produto e formas de interação; escolas e universidades decidem como integrar essas ferramentas ao ensino; serviços de saúde avaliam possibilidades e limites de uso; governos estabelecem normas e políticas públicas. Cada ator possui responsabilidades e capacidades distintas.

Nas instituições educacionais, o desafio não se resume a permitir ou proibir ferramentas. É necessário desenvolver letramento em IA, ensinar a verificar informações, discutir autoria e integridade acadêmica e preservar o papel da reflexão própria no processo de aprendizagem. O estudante precisa compreender que uma resposta bem estruturada pode conter erros e que a elaboração de um argumento exige avaliação das fontes, não apenas aceitação de um texto produzido por um sistema.

Nos serviços de saúde e no campo da Psicologia, a utilização de IA exige atenção adicional à privacidade, à segurança, aos limites de competência e à responsabilidade profissional. Sistemas podem apoiar tarefas informativas ou administrativas, mas não devem ser tratados automaticamente como substitutos do julgamento clínico, da relação terapêutica ou das obrigações éticas do profissional. O uso adequado depende do propósito, do contexto e das salvaguardas existentes.

No debate específico sobre psicoterapia, Ivantes-Rodrigues et al. (2025) examinam limites tecnológicos e desafios bioéticos associados ao uso da inteligência artificial na promoção da saúde. Os autores destacam questões como confidencialidade, privacidade, consentimento e os limites da tecnologia na construção de uma relação terapêutica. Essa contribuição reforça a necessidade de distinguir ferramentas de apoio de práticas que exigem vínculo profissional, julgamento clínico e responsabilidade ética.

Nas organizações públicas e privadas, é importante que decisões sobre adoção de IA incluam avaliação de riscos, transparência, proteção de dados, mecanismos de contestação e supervisão humana compatível com a gravidade das consequências. O documento do National Institute of Standards and Technology dedicado à IA generativa organiza riscos e práticas de gestão ao longo do desenvolvimento e da utilização dos sistemas (AUTIO et al., 2024). Embora seja uma referência técnica e não uma lei, oferece uma estrutura útil para pensar a responsabilidade institucional.

A sociedade também precisa considerar desigualdades de acesso e de capacidade de avaliação. Pessoas com diferentes níveis de letramento digital, escolaridade, recursos e familiaridade tecnológica podem enfrentar riscos distintos. Uma política de uso responsável não pode transferir toda a responsabilidade ao usuário e ignorar as condições de acesso à informação, os incentivos de mercado e as escolhas de projeto.

Desse modo, a pergunta sobre como as pessoas se relacionam com a IA deve ser acompanhada por outra: que condições sociais e institucionais estão sendo criadas para que essa relação seja informada, segura e compatível com os direitos humanos? Essa mudança de foco amplia a contribuição da Psicologia sem retirar o indivíduo do centro da análise.

6.1 Inteligência artificial na docência: autonomia, autoria e responsabilidade acadêmica

No contexto educacional, a presença da inteligência artificial exige uma reflexão que ultrapasse a discussão sobre sua permissão ou proibição em atividades acadêmicas. Para docentes e estudantes, o desafio consiste em compreender como essas ferramentas podem contribuir para a aprendizagem sem comprometer a autonomia intelectual, a autoria e a responsabilidade pela produção do conhecimento.

Na produção científica brasileira, Peixoto (2025) discute os impactos da IA generativa na formação e no exercício profissional da Psicologia e destaca a necessidade de competências críticas e éticas diante de sistemas que produzem linguagem e simulam interações. Essa reflexão também contribui para a formação de professores e estudantes: é preciso compreender as possibilidades das ferramentas, reconhecer seus limites e preservar a responsabilidade de quem aprende e ensina.

As normas da Associação Brasileira de Normas Técnicas (ABNT) oferecem referências para a adoção responsável da IA nas organizações. A ABNT NBR ISO/IEC 22989:2023 trata de conceitos e terminologia; a ABNT

NBR ISO/IEC 23894:2023 orienta a gestão de riscos; e a ABNT NBR ISO/IEC 42001:2024 estabelece requisitos para sistemas de gestão de IA. Essas normas não substituem decisões pedagógicas nem definem, isoladamente, regras universais para o uso em sala de aula, mas evidenciam a importância do planejamento, da avaliação de riscos e da definição de responsabilidades (OLIVEIRA et al., 2025).

Em outubro de 2026, o Conselho Federal de Psicologia, ao noticiar debate promovido pela Rede de Periódicos de Psicologia (PePsic), destacou a importância do uso responsável e transparente da IA na produção e na comunicação científica, com atenção à integridade acadêmica, à confiabilidade das informações e à preservação da autoria (CONSELHO FEDERAL DE PSICOLOGIA, 2026). A questão alcança diretamente a docência, pois estudantes e professores já utilizam essas ferramentas para pesquisar, organizar ideias e elaborar textos.

Diante desse cenário, cabe ao professor não apenas orientar sobre o uso das ferramentas, mas também criar condições para que os estudantes aprendam a questionar, verificar e avaliar criticamente os conteúdos produzidos. Uma resposta bem estruturada não é, necessariamente, uma resposta correta; assim como um texto fluente não comprova, por si só, compreensão do tema. A mediação docente permanece essencial para que a aprendizagem não se reduza à obtenção de respostas prontas.

Há, ainda, uma questão relacionada à autoria. Quando uma ferramenta reorganiza argumentos, reformula parágrafos ou sugere referências, pode contribuir para o processo de escrita, mas também pode modificar o sentido das ideias, omitir perspectivas relevantes ou apresentar informações incorretas. Por isso, o uso acadêmico da inteligência artificial exige revisão, conferência das fontes e transparência quanto à participação da ferramenta, de acordo com as orientações da instituição de ensino ou do periódico científico.

A questão central não é afastar a inteligência artificial do ambiente educacional, mas compreender quais usos favorecem a aprendizagem e quais podem enfraquecer o pensamento. O professor precisa reconhecer as possibilidades da tecnologia sem abdicar de sua responsabilidade pedagógica; o estudante, por sua vez, deve compreender que o apoio de uma ferramenta não o dispensa de entender, avaliar e responder pelo que apresenta como produção acadêmica. A autonomia não exige realizar todas as tarefas sem auxílio, mas conservar a capacidade de compreender, deliberar e assumir responsabilidade pelas próprias escolhas.

7 DIMENSÕES ÉTICAS E JURÍDICAS: BRASIL E REFERÊNCIAS INTERNACIONAIS

A discussão sobre autonomia, vínculos e confiança exige considerar os marcos jurídicos e éticos que orientam o desenvolvimento e o uso da inteligência artificial. É necessário, contudo, distinguir legislação vigente, projetos de lei, recomendações internacionais e estruturas voluntárias de gestão de riscos, pois cada instrumento possui natureza e força normativa diferentes.

No Brasil, a Lei nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais (LGPD), estabelece princípios e regras para o tratamento de dados pessoais, inclusive em meios digitais. Entre os princípios previstos estão finalidade, adequação, necessidade, transparência, prevenção, não discriminação e responsabilização (BRASIL, 2018). Esses princípios são relevantes para sistemas de IA que tratam dados pessoais, embora a LGPD não constitua, por si só, um marco regulatório completo para todos os aspectos da inteligência artificial.

Outro instrumento brasileiro a acompanhar é o Projeto de Lei nº 2.338/2023, voltado à regulação do desenvolvimento, do fomento e do uso ético e responsável da inteligência artificial. O projeto foi aprovado pelo Senado Federal em dezembro de 2024 e encaminhado à Câmara dos Deputados em março de 2025. Na consulta à ficha oficial de tramitação, em 2 de novembro de 2026, constava como aguardando parecer do relator na Comissão Especial. Portanto, ainda não deve ser apresentado como lei vigente (BRASIL, 2023). Como se trata de matéria em tramitação, sua situação deverá ser novamente conferida na data de submissão do artigo.

No plano internacional, a Recomendação da UNESCO sobre a Ética da Inteligência Artificial estabelece princípios relacionados à dignidade humana, aos direitos, à transparência, à supervisão e à avaliação dos impactos sociais (UNESCO, 2021). Por ser uma recomendação, não equivale automaticamente a uma lei nacional vinculante, mas oferece referência ética para orientar políticas e práticas.

A União Europeia aprovou o Regulamento (UE) 2024/1689, conhecido como AI Act, que adota uma abordagem baseada em riscos e prevê obrigações distintas conforme os sistemas e suas aplicações (UNIÃO EUROPEIA, 2024). Suas disposições entram em aplicação de forma escalonada, conforme os prazos previstos no próprio regulamento; por isso, sua incidência deve ser examinada de acordo com o sistema e a finalidade de uso considerados.

Esses instrumentos oferecem pontos de comparação, mas não devem ser tratados como equivalentes. A LGPD concentra-se no tratamento de dados pessoais; a Recomendação da UNESCO apresenta princípios éticos; o AI Act estabelece obrigações jurídicas para sistemas e usos abrangidos por seu âmbito de aplicação; e um projeto de lei brasileiro ainda depende do processo legislativo. Essa distinção é fundamental para uma análise jurídica rigorosa.

Uma questão relevante para o artigo é se os marcos existentes contemplam suficientemente as dimensões psicológicas das interações com sistemas capazes de simular acolhimento, intimidade ou reciprocidade. A resposta não deve ser presumida. Precisamos examinar as disposições aplicáveis e identificar o que está expressamente previsto, o que pode ser abordado por normas gerais e o que permanece como questão para debate regulatório e pesquisa.

A proteção da autonomia psicológica não depende apenas da existência de uma norma específica sobre vínculos emocionais com IA. Pode envolver, conforme o caso, proteção de dados, direitos do consumidor, normas profissionais, transparência, segurança e regras sobre práticas manipulativas. Ainda assim, é necessário investigar se esses instrumentos oferecem proteção suficiente diante das características particulares de sistemas de companhia e interação emocional.

8 DISCUSSÃO: COMO CONFIAR E COMO APRENDER A CONFIAR?

As dimensões psicológica, social, institucional e jurídica convergem para uma questão: como desenvolver formas de confiança que reconheçam tanto as possibilidades quanto os limites da inteligência artificial? A confiança facilita o uso, mas pode favorecer a aceitação de informações incorretas quando não é acompanhada de avaliação crítica. A percepção de acolhimento pode tornar a interação significativa para o usuário sem demonstrar que o sistema possui sentimentos ou compreensão subjetiva.

Uma contribuição central da Psicologia é examinar os processos pelos quais as pessoas interpretam sinais sociais, atribuem intenção e constroem expectativas. A fluência da linguagem, a continuidade do diálogo e a personalização podem ser interpretadas como sinais de competência ou compreensão. Entretanto, esses sinais não substituem a avaliação do desempenho real do sistema na tarefa em questão.

A confiança adequada deve ser calibrada. Para uma tarefa simples de organização de ideias, o risco pode ser relativamente baixo e a resposta pode ser avaliada com facilidade. Para uma decisão clínica, jurídica ou de segurança, as consequências de um erro podem ser maiores, exigindo verificação independente e participação humana qualificada. O nível de escrutínio deve acompanhar a importância da decisão e a possibilidade de dano.

Também é preciso considerar a dimensão relacional. Se o usuário percebe a IA como companhia ou fonte de apoio, a confiança pode envolver não apenas a crença na precisão das respostas, mas expectativas de disponibilidade, aceitação e continuidade. Esses fatores tornam necessário investigar como os sistemas são projetados e quais incentivos orientam sua utilização. A responsabilidade não pode recair exclusivamente sobre pessoas que talvez não conheçam os mecanismos técnicos e comerciais envolvidos.

Do ponto de vista institucional, a educação para o uso da IA precisa ultrapassar instruções sobre como escrever comandos. Deve incluir compreensão dos limites dos modelos, verificação de fontes, proteção de dados, reconhecimento de vieses e reflexão sobre a delegação de decisões. A literacia em IA também deve permitir que os usuários identifiquem quando uma resposta exige consulta a um profissional, a uma fonte primária ou a outra forma de avaliação.

A reflexão sobre a sociedade que utiliza IA para estudar a própria IA é especialmente significativa. Esses sistemas podem apoiar o processo de elaboração intelectual, mas não eliminam a responsabilidade de quem formula a pergunta, seleciona as fontes, interpreta as respostas e decide o que será publicado. O uso da ferramenta deve ser acompanhado de transparência metodológica e verificação das referências, especialmente em trabalhos acadêmicos.

Assim, o objetivo não é demonizar a tecnologia nem atribuir-lhe capacidades humanas que não foram demonstradas. O desafio é construir uma relação crítica e informada, na qual a IA possa ampliar possibilidades de aprendizagem e reflexão, sem substituir indiscriminadamente o julgamento humano ou obscurecer a responsabilidade de seus desenvolvedores e usuários.

9 IMPLICAÇÕES PARA A PSICOLOGIA E AGENDA DE PESQUISA

A Psicologia pode contribuir para a compreensão das interações humano-IA em diferentes níveis. No plano individual, pode investigar confiança percebida, antropomorfização, expectativas, regulação emocional e formas de delegação de decisões. No plano relacional, pode estudar a experiência de companhia, a percepção de reciprocidade e os limites entre apoio percebido e substituição de relações humanas. No plano institucional, pode contribuir para diretrizes de uso responsável, formação profissional e avaliação de riscos.

É importante que pesquisas futuras distingam tipos de sistemas e finalidades de uso. Um assistente de produtividade, um tutor educacional, uma ferramenta de apoio clínico e uma IA desenvolvida para companhia não possuem necessariamente os mesmos objetivos, interfaces ou riscos. Generalizar resultados entre esses contextos pode produzir conclusões imprecisas.

Também são necessárias pesquisas com diferentes grupos populacionais e desenhos metodológicos. Estudos qualitativos podem aprofundar o significado subjetivo das interações; estudos longitudinais podem examinar mudanças ao longo do tempo; pesquisas experimentais podem testar como características de interface afetam confiança e tomada de decisão; e revisões sistemáticas podem mapear a qualidade e os limites das evidências existentes.

Outra linha de investigação envolve o papel da transparência. É preciso examinar em que condições informar que o interlocutor é um sistema artificial altera a confiança, a percepção de acolhimento ou a forma como o usuário avalia a resposta. A transparência, embora importante, pode não ser suficiente se a interface continuar estimulando interpretações equivocadas ou se as limitações do sistema não forem comunicadas de maneira compreensível.

Para a prática profissional, a discussão sugere a importância de desenvolver competências para avaliar criticamente ferramentas de IA, proteger informações confidenciais e manter a responsabilidade humana em decisões relevantes. Essas orientações precisam ser articuladas às normas profissionais aplicáveis e atualizadas à medida que as ferramentas e as evidências científicas evoluem.

Uma agenda de pesquisa consistente deve evitar conclusões antecipadas sobre benefícios ou danos universais. O objetivo é identificar

condições, mecanismos e contextos em que as interações com IA apoiam a autonomia ou podem comprometer a avaliação crítica, o bem-estar e a qualidade dos vínculos.

10 CONSIDERAÇÕES FINAIS

A inteligência artificial amplia as possibilidades de interação entre pessoas e sistemas computacionais e cria questões importantes para a Psicologia e para as instituições responsáveis pela proteção de direitos. Autonomia, vínculos e confiança constituem dimensões fundamentais para compreender essas relações, sobretudo quando a comunicação artificial assume características percebidas como acolhedoras, personalizadas ou semelhantes às interações humanas.

A confiança responsável depende do reconhecimento das capacidades e limitações dos sistemas. Uma resposta fluente não é garantia de verdade, assim como a percepção de acolhimento não demonstra a existência de sentimentos ou consciência no sistema. A autonomia exige que as pessoas mantenham a capacidade de avaliar informações, questionar respostas e decidir sobre suas próprias ações, utilizando a IA como recurso de apoio quando apropriado.

REFERENCIAS BIBLIOGRAFICAS:

AMERICAN PSYCHOLOGICAL ASSOCIATION (APA). Artificial intelligence and the field of psychology. Washington, DC: American Psychological Association, 2024. Disponível em: https://www.apa.org/about/policy/statement-artificial-intelligence.pdf. Acesso em: 2 nov. 2026.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO/IEC 22989:2023: tecnologia da informação — inteligência artificial — conceitos de inteligência artificial e terminologia. Rio de Janeiro: ABNT, 2023.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO/IEC 23894:2023: tecnologia da informação — inteligência artificial — orientações sobre gestão de riscos. Rio de Janeiro: ABNT, 2023.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO/IEC 42001:2024: tecnologia da informação — inteligência artificial — sistema de gestão. Rio de Janeiro: ABNT, 2024.

AUTIO, Chloe et al. Artificial intelligence risk management framework: generative artificial intelligence profile. Gaithersburg, MD: National Institute of Standards and Technology, 2024. (NIST AI 600-1). DOI: https://doi.org/10.6028/NIST.AI.600-1. Disponível em: https://doi.org/10.6028/NIST.AI.600-1. Acesso em: 25 out. 2026.

BENDER, Emily M. et al. On the dangers of stochastic parrots: can language models be too big? In: ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021, Virtual Event. Proceedings [...]. New York: Association for Computing Machinery, 2021. p. 610–623. DOI: https://doi.org/10.1145/3442188.3445922. Disponível em: https://doi.org/10.1145/3442188.3445922. Acesso em: 25 out. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 2 nov. 2026.

BRASIL. Câmara dos Deputados. Projeto de Lei nº 2.338, de 2023. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF: Câmara dos Deputados, 2023. Situação da tramitação consultada em: 2 nov. 2026. Disponível em: https://www.camara.leg.br/proposicoesWeb/fichadetramitacao/?idProposicao=2487262.

CONSELHO FEDERAL DE PSICOLOGIA. Inteligência artificial na Psicologia: guia para uma prática ética e responsável. Brasília, DF: CFP, 2025. Disponível em: https://site.cfp.org.br/wp-content/uploads/2025/12/Cartilha_IA_A5-1.pdf. Acesso em: 2 nov. 2026.

CONSELHO FEDERAL DE PSICOLOGIA. Rede PePsic promove debate sobre uso de inteligência artificial no contexto científico. Brasília, DF, 2 out. 2026. Disponível em: https://site.cfp.org.br/pepsic2026/. Acesso em: 2 nov. 2026.

IVANTES-RODRIGUES, Sophia; SILVA, Marilia da Matta; OLIVEIRA, Leonardo Pestillo de; GARCIA, Lucas França. Psicoterapia e inteligência artificial: limites tecnológicos na promoção da saúde. Revista Bioética, Brasília, DF, v. 33, e3958PT, 2025. DOI: https://doi.org/10.1590/1983-803420253958PT. Disponível em: https://www.scielo.br/j/bioet/a/RGvGPb5FY8PgwHkkjGkRGCd/?lang=pt. Acesso em: 27 out. 2026.

LEE, John D.; SEE, Katrina A. Trust in automation: designing for appropriate reliance. Human Factors, v. 46, n. 1, p. 50–80, 2004. DOI: https://doi.org/10.1518/hfes.46.1.50_30392. Disponível em: https://user.engineering.uiowa.edu/~csl/publications/pdf/leesee04.pdf. Acesso em: 27 out. 2026.

NASS, Clifford; MOON, Youngme. Machines and mindlessness: social responses to computers. Journal of Social Issues, v. 56, n. 1, p. 81–103, 2000. DOI: https://doi.org/10.1111/0022-4537.00153. Disponível em: https://spssi.onlinelibrary.wiley.com/doi/pdf/10.1111/0022-4537.00153. Acesso em: 25 out. 2026.

OLIVEIRA, Danielle Teixeira de; ALMEIDA, Aline Pimenta; SOARES, Jordana Rabelo; NASCIMENTO SILVA, Patrícia; OLIVEIRA, Dalgiza Andrade. Normas regulamentadoras da inteligência artificial no Brasil: definições da ABNT. Revista Conhecimento em Ação, Rio de Janeiro, v. 10, e65554, 2025. DOI: https://doi.org/10.47681/rca.v10i.65554. Disponível em: https://revistas.ufrj.br/index.php/rca/article/view/65554. Acesso em: 2 nov. 2026.

PEIXOTO, Adriano de Lemos Alves. Inteligência artificial generativa e seu impacto na formação e no exercício profissional da Psicologia. Psicologia: Ciência e Profissão, Brasília, DF, v. 45, n. spe1, e298224, 2025. DOI: https://doi.org/10.1590/1982-3703003298224. Disponível em: https://www.scielo.br/j/pcp/a/p3TgdLYJJTkLhzYy8K5T7wf/?lang=pt. Acesso em: 2 nov. 2026.

PITARDI, Valentina; MARRIOTT, Hannah R. Alexa, she's not human but… Unveiling the drivers of consumers' trust in voice-based artificial intelligence. Psychology & Marketing, v. 38, n. 4, p. 626–642, 2021. DOI: https://doi.org/10.1002/mar.21457. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/mar.21457. Acesso em: 27 out. 2026.

SARACINI, Chiara; CORNEJO-PLAZA, María Isabel; CIPPITANI, Roberto. Techno-emotional projection in human–GenAI relationships: a psychological and ethical conceptual perspective. Frontiers in Psychology, v. 16, art. 1662206, 2025. DOI: https://doi.org/10.3389/fpsyg.2025.1662206. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12515930/. Acesso em: 25 out. 2026.

SHANAHAN, Murray. Talking about large language models. Communications of the ACM, v. 67, n. 2, p. 68–79, 2024. DOI: https://doi.org/10.1145/3624724. Disponível em: https://doi.org/10.1145/3624724. Acesso em: 2 nov. 2026.

UNESCO. Recommendation on the ethics of artificial intelligence. Paris: UNESCO, 2021. Disponível em: https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence. Acesso em: 27 out. 2026.

UNIÃO EUROPEIA. Parlamento Europeu; Conselho da União Europeia. Regulamento (UE) 2024/1689, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial (Regulamento da Inteligência Artificial). Jornal Oficial da União Europeia, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 25 out. 2026.

O princípio da proteção do trabalhador frente à inteligência artificial: limites ao poder dos algoritmos

The principle of worker protection against artificial intelligence: limits to the power of algorithms.

*Fabiane Regina Carvalho de Andrade Ibrahin**

RESUMO

Este capítulo examina como o princípio da proteção do trabalhador pode orientar os limites ao uso da inteligência artificial na definição de metas e na avaliação de desempenho. Parte-se da hipótese de que o princípio, lido como critério de limitação do poder e não apenas como regra de interpretação, exige que a dignidade do trabalhador limite o poder diretivo, que esse limite alcance os meios tecnológicos de seu exercício e que quem detém o poder e a informação demonstre a legitimidade de suas decisões. A pesquisa é jurídico-dogmática e qualitativa, centrada no direito brasileiro, com análise de normas, doutrina e decisões judiciais. Examinam-se os riscos do gerenciamento algorítmico, como metas inalcançáveis, rankings contínuos, discriminação indireta e adoecimento psíquico, e os limites oferecidos pela Lei Geral de Proteção de Dados, pela Norma Regulamentadora nº 1 e pelo Projeto de Lei nº 2.338/2023. Conclui-se que essas normas protegem apenas parcialmente e propõem-se critérios materiais e procedimentais para o uso responsável da IA no trabalho, além da responsabilização do empregador.

Palavras-chave: Princípio da proteção. Inteligência artificial. Poder diretivo. Gestão algorítmica. Avaliação de desempenho.

---

* Pós doutoranda pelo Programa Interunidades Integração da América Latina PROLAM/USP, Doutora em Ciências Sociais pela Pontifícia Universidade Católica de São Paulo - PUC. Mestre em Direito Político e Econômico pela Universidade Presbiteriana Mackenzie. Mestre em Direito pela UNIFIEO. Especialista em Direito do Trabalho pela UNIFIEO. Graduada em Direito pela ITE - Instituição Toledo de Ensino/Araçatuba. Pesquisadora do GETRAB/USP - Grupo de Estudos de Direito Contemporâneo do Trabalho e da Seguridade Social. Pesquisadora do CIIDCMIL/USP - Centro Internacional de Inovação e Desenvolvimento de Cidades MIL. Professora Universitária. Coordenadora do Curso de Direito da Faculdade Lusófona/SP. Escritora. Palestrante. Avaliadora Institucional do INEP/MEC.

ABSTRACT

This chapter examines how the principle of worker protection can guide limits on the use of artificial intelligence in goal setting and performance evaluation. It starts from the hypothesis that the principle, read as a criterion for limiting power rather than merely a rule of interpretation, requires that workers' dignity limit managerial power, that this limit extend to the technological means of exercising it, and that those who hold power and information demonstrate the legitimacy of their decisions. The research is doctrinal and qualitative, focused on Brazilian law, and analyzes legislation, scholarship and court decisions. It examines the risks of algorithmic management, such as unattainable targets, continuous rankings, indirect discrimination and psychological illness, and the limits offered by the General Data Protection Law, Regulatory Standard No. 1 and Bill No. 2,338/2023. It concludes that these rules offer only partial protection and proposes substantive and procedural criteria for the responsible use of AI at work, together with employer accountability.

Keywords: Principle of protection. Artificial intelligence. Managerial power. Algorithmic management. Performance evaluation.

1 INTRODUÇÃO

A inteligência artificial deixou de ser apenas ferramenta de apoio e passou a participar de decisões que organizam a vida em sociedade, como a concessão de crédito, a exibição de conteúdo e a seleção para uma vaga. Com isso, transforma as relações sociais em dois sentidos: desloca decisões antes tomadas entre pessoas, sujeitas a justificação e contestação, para sistemas que operam de forma contínua e pouco compreensível para quem é afetado; e amplia a assimetria de informação entre quem projeta e utiliza o sistema e quem se submete a ele. A pergunta que orienta esta obra, sobre como a IA transforma as relações sociais e quais critérios devem orientar seu uso responsável, encontra no trabalho um terreno particularmente nítido.

No trabalho subordinado, uma das partes dirige, fiscaliza e pune a outra, e a dependência econômica limita a capacidade de resistir. Quando metas e avaliações passam a ser calculadas por sistemas automatizados, o poder de direção, que sempre teve limites jurídicos, ganha novos meios: monitoramento contínuo, comparação permanente e decisão sem rosto. O problema jurídico está em que o poder diretivo é funcional e limitado, e sua legitimidade depende do respeito à dignidade do trabalhador, ao passo que o gerenciamento algorítmico dificulta a verificação desse limite, porque os critérios das metas e das notas raramente são conhecidos por quem é avaliado. O ordenamento brasileiro dispõe de normas aplicáveis, como a LGPD e a NR-1, mas nenhuma foi concebida para esse objeto, e o projeto de lei que o enfrenta ainda não foi aprovado.

Nesse contexto, pergunta-se: como o princípio da proteção do trabalhador pode orientar os limites ao uso da inteligência artificial na definição de metas e na avaliação de desempenho? Sustenta-se que o princípio, entendido como critério de limitação do poder e não apenas como regra de interpretação, permite extrair exigências materiais e procedimentais para o uso dos sistemas. O objetivo é identificar esses critérios e relacioná-los aos instrumentos jurídicos existentes e às lacunas que persistem.

A pesquisa é jurídico-dogmática, de natureza qualitativa, centrada no direito brasileiro, com análise de normas, doutrina e decisões judiciais. Como não foram localizadas decisões trabalhistas brasileiras sobre metas definidas por algoritmo, a jurisprudência sobre cobrança abusiva de metas é usada como parâmetro por analogia, e os efeitos próprios dos sistemas automatizados são tratados como hipóteses. O capítulo examina o princípio da proteção e o poder diretivo na era algorítmica (seção 2), os riscos das metas e da avaliação por IA (seção 3) e os limites jurídicos existentes e suas lacunas (seção 4); em seguida, propõe critérios para o uso responsável da IA na gestão do trabalho (seção 5), antes das considerações finais (seção 6).

2 O PRINCÍPIO DA PROTEÇÃO E O PODER DIRETIVO NA ERA ALGORÍTMICA

Antes de examinar como a inteligência artificial interfere em metas e avaliações de desempenho, é preciso estabelecer o parâmetro de julgamento dessa interferência. Este capítulo o constrói em três passos: reconstrói o princípio da proteção e sua base constitucional, para defender que ele é critério de limitação do poder; examina o poder diretivo como poder funcional, sujeito a limites; e descreve como o gerenciamento algorítmico altera as condições em que esse poder é exercido e controlado. Ao final, fixam-se as premissas que orientam a pesquisa.

2.1 O Princípio da proteção e seus fundamentos constitucionais

A doutrina trabalhista costuma tomar o princípio da proteção como instrumento de interpretação e de solução de conflitos entre normas (PLÁ RODRIGUEZ, 2000). A leitura é correta, mas insuficiente. Metas e avaliações definidas por sistemas de inteligência artificial não geram, em regra, dúvida sobre o sentido de um texto normativo: geram decisões, muitas vezes

sem texto a interpretar e sem uma pessoa identificável que as justifique. Para esse cenário, o princípio deve ser lido não apenas como regra de hermenêutica, mas como critério de limitação do poder.

A razão está na estrutura da relação de emprego. O trabalhador se subordina porque depende do salário, e o empregador exerce poder porque detém os meios de organizar a atividade. A proteção existe para que essa assimetria, inerente ao contrato, não se converta em domínio (PLÁ RODRIGUEZ, 2000). O poder diretivo é, assim, reconhecido e condicionado.

A Constituição de 1988 transformou essa lógica em compromisso explícito. Coloca entre os fundamentos da República "a dignidade da pessoa humana" e "os valores sociais do trabalho e da livre iniciativa" (BRASIL, 1988, art. 1º, III e IV) e funda a ordem econômica "na valorização do trabalho humano e na livre iniciativa", para "assegurar a todos, existência digna, conforme os ditames da justiça social" (art. 170, caput). O art. 7º enumera direitos "além de outros que visem à melhoria de sua condição social", o que faz do patamar constitucional um piso, e não um teto. Livre iniciativa e valor social do trabalho aparecem lado a lado, sem hierarquia: a primeira legitima o empregador a organizar a produção, fixar objetivos e medir resultados; a dignidade e o valor social do trabalho definem até onde o poder pode ir.

O limite atinge também a integridade psíquica. A Constituição prevê a "redução dos riscos inerentes ao trabalho, por meio de normas de saúde, higiene e segurança" (BRASIL, 1988, art. 7º, XXII) e inclui o meio ambiente do trabalho na proteção ambiental (art. 200, VIII). Isso é decisivo, porque a gestão algorítmica de metas modifica menos o salário ou o emprego em si e mais o ritmo, a pressão e a exposição contínua de quem trabalha; sua dimensão psicossocial integra, portanto, o objeto da proteção (seção 3.3).

A Constituição também assegura a "proteção em face da automação, na forma da lei" (BRASIL, 1988, art. 7º, XXVII). Em 2025, ao reconhecer a omissão do Congresso Nacional em regulamentá-la e fixar prazo de 24 meses, o Supremo Tribunal Federal destacou que a proteção deve ser conciliada com a livre iniciativa e com o incentivo ao desenvolvimento tecnológico (BRASIL, 2025). Automação, porém, designa sobretudo a substituição de postos de trabalho por máquinas, e não necessariamente a gestão do trabalho por algoritmos; por isso o inciso é tratado aqui como reforço, e não como fundamento exclusivo.

Disso se extraem três premissas. A primeira: o poder do empregador encontra limite na dignidade do trabalhador. A segunda: esse limite alcança os meios tecnológicos pelos quais o poder é exercido, de modo que delegar a decisão a um sistema não suspende a proteção. A terceira: quem detém

o poder e a informação sobre o funcionamento do sistema deve estar em condições de demonstrar a legitimidade de suas decisões. Elas orientam o restante do capítulo e fundamentam os critérios da seção 5.

2.2 O poder diretivo como poder funcional e limitado

O poder diretivo é a face prática da subordinação: por meio dele o empregador organiza a atividade, distribui tarefas, fixa objetivos e controla resultados. A doutrina o situa dentro de um poder empregatício mais amplo, que comporta também as dimensões regulamentar, fiscalizatória e disciplinar (DELGADO, 2019). Interessa aqui menos catalogá-las e mais perguntar a que serve o poder e onde ele termina.

A primeira resposta vem do conceito legal de empregador: quem, "assumindo os riscos da atividade econômica, admite, assalaria e dirige a prestação pessoal de serviço" (BRASIL, 1943, art. 2º). A lei põe lado a lado o poder de dirigir e o dever de arcar com os riscos do empreendimento. O poder de comando é reconhecido porque o empregador responde pelo resultado da atividade. Daí um critério de funcionalidade: o poder diretivo serve para organizar a produção, e não para deslocar ao trabalhador, por metas calibradas ao limite, o risco que a lei atribuiu a quem empreende. Fora dessa finalidade, deixa de ser poder legítimo e passa a ser abuso.

Esse limite finalístico se soma à lei, ao contrato, às normas coletivas e à Constituição. O Código Civil considera ilícito o exercício de um direito que "excede manifestamente os limites impostos pelo seu fim econômico ou social, pela boa-fé ou pelos bons costumes" (BRASIL, 2002, art. 187) e exige a boa-fé na conclusão e na execução do contrato (art. 422). A CLT, ao admitir a rescisão quando exigidos "serviços superiores às suas forças" ou quando houver "rigor excessivo" (BRASIL, 1943, art. 483, a e b), reconhece desde a origem que o comando pode ser exercido de forma excessiva.

O art. 6º, parágrafo único, da CLT, na redação da Lei nº 12.551/2011, equipara os meios telemáticos e informatizados de comando, controle e supervisão aos meios pessoais e diretos, para fins de subordinação jurídica (BRASIL, 1943, art. 6º, parágrafo único). O legislador reconheceu, assim, que o poder do empregador pode ser exercido por sistemas e os tratou como extensão desse poder, e não como terceiros neutros. A norma foi concebida para caracterizar a subordinação, e não para fixar limites; defende-se aqui, porém, uma leitura teleológica: se o comando algorítmico equivale ao pessoal para fundar a subordinação, deve equivaler também para sujeitar-se aos mesmos limites. Não seria coerente atribuir ao sistema o poder do empregador, mas não os deveres que o acompanham.

A boa-fé objetiva impõe ainda deveres de informação e transparência. Sem saber que comportamentos são medidos e como pesam na avaliação, o empregado não ajusta sua conduta, não questiona um resultado nem se defende de uma sanção. Por isso a legitimidade do poder exercido por algoritmo deve poder ser demonstrada por quem o exerce, pois é o empregador que conhece o sistema.

Em síntese, o poder diretivo é funcional, vinculado à organização da atividade que o empregador assume por sua conta e risco, e limitado pela lei, pelo contrato, pela boa-fé e pela dignidade do trabalhador. Delegar seu exercício a um sistema muda a forma como ele se manifesta, e não sua natureza.

2.3 Do chefe humano ao gerenciamento algorítmico

o gerenciamento algorítmico caracteriza-se pelo uso de sistemas automatizados em funções tradicionalmente desempenhadas pelo empregador, como monitoramento, vigilância, avaliação de desempenho e decisões sobre a organização e a execução do trabalho. Conforme Rodrigues (2026), representa a automação de mecanismos de gestão e controle antes exercidos diretamente pelo empregador. Na sistematização de Kellogg, Valentine e Christin (2020), esses sistemas dirigem o trabalho, ao recomendar e restringir condutas; avaliam, ao registrar e classificar o desempenho; e disciplinam, ao premiar ou substituir quem não atinge o esperado.

O fenômeno ganhou destaque nos estudos sobre plataformas digitais, nas quais Kalil (2020) demonstra diferentes intensidades de controle, com repercussões sobre a autonomia e as condições de trabalho. Seria equívoco, porém, tê-lo como exclusivo das plataformas. Conforme Miziara (2024), ferramentas de IA podem ser empregadas em diversas etapas da relação de emprego, do recrutamento à gestão contratual, incluindo decisões sobre promoção e dispensa, e também podem monitorar indicadores de produtividade, avaliar o desempenho individual e subsidiar a definição de metas. O objeto deste capítulo abrange, portanto, as relações de emprego em sentido amplo.

É tentador contrapor o chefe humano, falível e subjetivo, ao algoritmo, objetivo e imparcial. A comparação está mal colocada: gestores humanos também são parciais, e muitos abusos registrados na jurisprudência foram praticados por pessoas. A diferença relevante está nos mecanismos que, na gestão humana, moderavam o exercício do comando e que, na gestão algorítmica, deixam de existir, a menos que sejam deliberadamente incorporados ao sistema. Quatro deslocamentos merecem destaque.

O primeiro é a continuidade: o gestor observa em momentos determinados, e o sistema registra de modo contínuo e granular, transformando cada minuto em dado avaliável. O segundo é a opacidade: o empregado pode perguntar ao chefe o motivo de uma cobrança e apresentar o contexto de um mau resultado, mas diante do sistema muitas vezes não sabe sequer quais comportamentos são medidos e como pesam na nota. O terceiro é a assimetria informacional: quem projeta e opera o sistema conhece seus critérios, e quem é avaliado não, o que aprofunda a desigualdade que o princípio da proteção busca compensar. O quarto é a aparência de neutralidade: o número parece dispensar justificação, e a decisão pode ser atribuída ao sistema ("foi o algoritmo"), diluindo a responsabilidade de quem decidiu empregá-lo.

Esses deslocamentos explicam por que a gestão algorítmica de metas e avaliações exige atenção própria do Direito do Trabalho. Não se supõe que o sistema seja necessariamente pior, mas que altera as condições em que o poder diretivo é exercido e controlado. Se o comando por sistemas se equipara ao pessoal e deve respeitar os mesmos limites, é preciso identificar onde seu exercício produz riscos concretos ao trabalhador.

3 METAS E AVALIAÇÃO DE DESEMPENHO POR IA: RISCOS AO TRABALHADOR

A incorporação de sistemas de inteligência artificial à gestão do trabalho, especialmente na definição de metas e na avaliação de desempenho, pode contribuir para a organização e a eficiência das atividades, mas também intensificar o controle e ampliar situações de vulnerabilidade. A análise volta-se a três riscos: a imposição e a cobrança abusiva de metas; as práticas discriminatórias decorrentes de classificações e vieses algorítmicos; e os impactos à saúde física e mental pela pressão contínua por resultados.

3.1 Metas dinâmicas e cobrança abusiva

Fixar metas é parte legítima do poder diretivo. O Tribunal Superior do Trabalho já afirmou que definir e cobrar metas integra esse poder, de modo que exigir patamar mínimo de resultados não é, em si, ilícito; a responsabilidade surge quando o empregador ultrapassa seus limites e submete o empregado a situações ofensivas (BRASIL, 2015, RR-528-74.2011.5.09.0001, 2ª Turma). Em outro precedente, reconheceu que a cobrança de metas que ultrapassa a razoabilidade e afronta a dignidade

da pessoa humana pode configurar assédio moral, caracterizado pela prática reiterada de ofensas, castigos e situações humilhantes pelo não cumprimento das metas (BRASIL, 2017b, RR-68300-89.2009.5.09.0012, 4ª Turma). Essa jurisprudência não nasceu da gestão algorítmica, mas fornece o parâmetro que este capítulo propõe estender a ela.

Dos precedentes extraem-se três critérios. O primeiro é a viabilidade das metas: o Tribunal reconheceu assédio moral em caso de gerente bancária submetida a metas inviáveis, mantidas após a redução da equipe, sem adequação das exigências nem reposição de pessoal em prazo razoável, com cobranças depreciativas e agravamento de seu quadro de saúde (BRASIL, 2017a, RR-1485-42.2010.5.09.0088, 3ª Turma). O segundo é a forma da cobrança: em caso de coordenadora submetida a cerca de 540 ligações diárias, manteve-se a condenação por assédio moral diante de cobrança abusiva reiterada, exposição pública do desempenho em rankings e ameaças veladas de dispensa (BRASIL, 2026, RRAg-12520-13.2016.5.15.0026, 8ª Turma). O terceiro é o efeito sobre a saúde, pois a conexão entre a pressão por resultados e o agravamento de quadro depressivo foi decisiva na hipótese da bancária.

Cada critério tem correspondente na gestão algorítmica, e aí o risco se agrava. Sistemas podem recalcular metas continuamente a partir de dados agregados, sem considerar equipe reduzida, afastamentos ou variações de demanda; fixar metas relativas, em que alguém sempre ocupa a última posição, qualquer que seja o desempenho absoluto; elevar o patamar a cada ciclo tomando por referência o melhor resultado anterior, em escalada sem teto; e exibir rankings e alertas em tempo real, transformando em rotina automática a exposição que, nos casos acima, dependia de reuniões e mensagens. São possibilidades de configuração, e não afirmações sobre empresas específicas: nenhuma é ilícita por si, mas todas tensionam os critérios da viabilidade, da forma e da saúde.

Há ainda diferença na prova e na responsabilidade. Nos precedentes, o abuso foi demonstrado por testemunhas e documentos e atribuído a pessoas identificáveis. Diante de um sistema, a autoria se dilui, e as informações necessárias (critérios, parâmetros, histórico de ajustes) estão com o empregador. Daí a relevância das premissas da seção 2: o comando algorítmico é comando do empregador (BRASIL, 1943, art. 6º, parágrafo único), e a ele cabe demonstrar a legitimidade das metas que o sistema produz. Os rankings, os vieses e seus efeitos discriminatórios assumem outra feição, examinada a seguir.

3.2 Rankings, vieses e discriminação algorítmica

A ideia de que o algoritmo é neutro, por decidir com base em dados, não resiste à literatura sobre o tema. Barocas e Selbst (2016) mostram que sistemas de mineração de dados podem herdar os preconceitos de decisores anteriores e reproduzir padrões de exclusão, e que a discriminação costuma ser efeito não intencional do uso do sistema, o que dificulta identificar sua origem e explicá-la em juízo. Um caso emblemático foi o da ferramenta experimental de recrutamento da Amazon, treinada com currículos recebidos ao longo de cerca de dez anos, predominantemente de homens, que passou a penalizar candidaturas com referências associadas a mulheres (DASTIN, 2018). Trata-se de um caso de seleção, e não de avaliação de desempenho, mas o mecanismo é o mesmo: sempre que um sistema aprende o que é "bom desempenho" a partir de decisões e dados passados, pode reproduzir os vieses que os marcaram. A extensão desse raciocínio à avaliação de desempenho é uma hipótese deste capítulo.

A literatura trabalhista brasileira converge nesse diagnóstico. Teixeira, Pimentel e Pugliesi (2025) sustentam que os padrões algorítmicos podem produzir uma nova espécie de dano, a discriminação algorítmica, e Santos e Graminho (2024) defendem que os princípios da não discriminação, da transparência e da responsabilização e prestação de contas da Lei Geral de Proteção de Dados Pessoais (LGPD) ajudam a preveni-la. O ordenamento já veda a discriminação: a Constituição inclui entre os objetivos da República a promoção do bem de todos, sem preconceitos (BRASIL, 1988, art. 3º, IV) e proíbe diferenças de critérios de admissão por motivo de sexo, idade, cor ou estado civil (BRASIL, 1988, art. 7º, XXX); a Lei nº 9.029/1995 proíbe práticas discriminatórias para o acesso à relação de trabalho ou sua manutenção, abrangendo, entre outros fatores, sexo, origem, raça, cor, estado civil, situação familiar, deficiência e idade (BRASIL, 1995, art. 1º); e a LGPD impede o tratamento para fins discriminatórios ilícitos ou abusivos (BRASIL, 2018, art. 6º, IX).

Os rankings de desempenho mostram como é difícil delimitar o poder diretivo diante de entendimentos divergentes. Em 2026, a 6ª Turma do Tribunal Superior do Trabalho considerou abusiva a divulgação, por e-mail, de rankings de produtividade, por ultrapassar os limites do poder diretivo e configurar dano moral, independentemente de prova específica do prejuízo (BRASIL, 2026, RR-1001166-22.2016.5.02.0030, 6ª Turma). Em sentido diverso, a 2ª Turma do Tribunal Regional do Trabalho da 18ª Região entendeu que a divulgação interna de rankings não configura, por si só,

lesão à honra ou à dignidade, diante da ausência de tratamento humilhante ou abusivo e da existência de políticas de compliance e de prevenção ao assédio (BRASIL, 2025, ROT-0010553-89.2024.5.18.0018, 2ª Turma).

Os julgados permitem identificar elementos para delimitar o poder diretivo: a forma e a frequência da exposição dos resultados, o contexto da divulgação e a finalidade dos mecanismos de classificação. Na gestão algorítmica, esses fatores ganham nova dimensão, pois os rankings podem ser produzidos automaticamente, atualizados de modo contínuo e baseados em classificações relativas, que necessariamente colocam alguns trabalhadores nas últimas posições, ainda que atendam às metas individuais.

O risco discriminatório não depende do uso expresso de características protegidas. Métricas aparentemente neutras, como assiduidade, disponibilidade ou velocidade de resposta, podem prejudicar desproporcionalmente quem precisa de licença-maternidade, tratamento de saúde ou cuidado familiar, ou tem deficiência. Trata-se de possível discriminação indireta, que dispensa prova de intenção e exige a análise dos efeitos de critérios aparentemente neutros. Se a discriminação resulta no rompimento do contrato, a Lei nº 9.029/1995 assegura, além da reparação por dano moral, a opção entre a reintegração, com ressarcimento integral do período de afastamento, e a percepção em dobro da remuneração desse período, com correção monetária e juros legais (BRASIL, 1995, art. 4º).

Resta o problema da prova. O trabalhador dificilmente demonstra que uma nota ou meta decorreu de viés sem conhecer as variáveis e os dados do sistema. A LGPD prevê que, se o controlador deixar de informar os critérios da decisão automatizada com base em segredo comercial ou industrial, a autoridade nacional poderá realizar auditoria para verificar aspectos discriminatórios (BRASIL, 2018, art. 20, § 2º). Atribui, porém, a auditoria à autoridade, e não ao trabalhador que, em seu litígio, precisa da informação. A lacuna reforça a terceira premissa: cabe a quem detém o sistema demonstrar a legitimidade de suas decisões. A pressão contínua por resultados produz ainda outro dano, que independe de discriminação e é examinado a seguir.

3.3 Saúde mental e riscos psicossociais

A Constituição assegura a redução dos riscos inerentes ao trabalho e inclui o meio ambiente laboral na proteção ambiental (BRASIL, 1988, arts. 7º, XXII, e 200, VIII). No plano regulamentar, a Norma Regulamentadora

nº 1 (NR-1), na redação da Portaria MTE nº 1.419/2024, passou a exigir, a partir de 26 de maio de 2026, que o gerenciamento de riscos ocupacionais contemple os fatores de risco psicossociais relacionados ao trabalho (item 1.5.3.1.4), após a prorrogação promovida pela Portaria MTE nº 765/2025 (BRASIL, 2024, item 1.5.3.1.4; BRASIL, 2025). A literatura inclui entre esses fatores as metas desproporcionais, o assédio e a sobrecarga, e adverte que o trabalho, por si só, não adoece: a materialização do risco depende da intensidade, da duração e do contexto (PINTO JUNIOR; BOMFIM, 2026).

A gestão algorítmica de metas e avaliações incide sobre esses elementos. Segundo Bowdler et al. (2026), reconfigura o ambiente psicossocial ao transferir funções decisórias de gestores humanos para sistemas, o que pode intensificar as exigências e ampliar os riscos, sobretudo com pressão temporal, monitoramento rigoroso, redução de autonomia e falta de transparência dos critérios. Cefaliello, Moore e Donoghue (2023) acrescentam que metas definidas por algoritmos e vigilância contínua em tempo real geram sensação de controle permanente e ansiedade, e que a avaliação de riscos deve acompanhar o uso do sistema, e não apenas o seu desenho. Pesquisa com mais de seis mil trabalhadores nos países nórdicos encontrou mais pressão de tempo e sobrecarga, e menos autonomia e apoio social, onde o uso desses sistemas é mais intenso (TTL, 2025). São dados europeus, e os próprios pesquisadores ressaltam que o fator decisivo não é a tecnologia, mas a forma de implantação, com participação dos trabalhadores e transparência. Por isso se fala em risco, e não em efeito inevitável.

Quando o risco se concretiza, o dano é concreto. O TRT da 2ª Região reconheceu a síndrome de esgotamento profissional de uma gerente bancária como doença do trabalho (Lei nº 8.213/1991, arts. 20 e 21), ainda que o trabalho não fosse a única causa, e afastou a conclusão pericial em contrário por entender que a lei não exige que a condição seja catalogada como doença pela Organização Mundial da Saúde. Considerou ineficazes as medidas preventivas do empregador, como cursos e treinamentos, e fixou indenização por dano moral de R$ 50 mil, presumido, além de pensão mensal vitalícia (BRASIL, 2026, ROT-1000485-78.2025.5.02.0081, 1ª Turma). O caso não envolve algoritmo, e o acórdão não examinou a abusividade das metas, tendo rejeitado o pedido de indenização por assédio moral por ausência de prova. Mostra, porém, que medidas preventivas meramente formais podem ser insuficientes diante do adoecimento relacionado ao trabalho, o que reforça a necessidade de tratar os sistemas de metas e avaliação como fonte de risco.

A NR-1 exige o gerenciamento dos fatores de risco psicossociais, mas não disciplina expressamente os sistemas algorítmicos de metas e avaliação. Isso não afasta a incidência de suas disposições gerais quando o uso desses sistemas produzir riscos psicossociais relacionados à organização do trabalho (BRASIL, 2024, itens 1.5.3.1.4 e 1.5.3.2). Encerrada a análise dos riscos, cabe examinar o que o ordenamento já oferece para limitá-los e o que ainda falta.

4 LIMITES JURÍDICOS EXISTENTES E LACUNAS

O ordenamento brasileiro não é indiferente ao uso de sistemas automatizados na gestão do trabalho: há normas de proteção de dados, de saúde e segurança e um projeto de lei em tramitação que alcançam, em alguma medida, metas e avaliação de desempenho. Nenhuma, porém, foi concebida para esse objeto.

4.1 A lei geral de proteção de dados e a revisão de decisões automatizadas

A Lei nº 13.709/2018 reconhece ao titular o direito de solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, inclusive as destinadas a definir seu perfil profissional (BRASIL, 2018, art. 20). O controlador deve fornecer, quando solicitado, informações claras e adequadas sobre os critérios e procedimentos, observados os segredos comercial e industrial (§ 1º); se negar a informação com esse fundamento, a Autoridade Nacional de Proteção de Dados (ANPD) poderá realizar auditoria (§ 2º).

O dispositivo é a base mais próxima de uma garantia contra a avaliação algorítmica, mas tem três limites. O primeiro é a palavra “unicamente”: o direito não alcança o sistema que apenas produz a nota ou o ranking, quando um gestor o referenda, arranjo comum e que favorece a aparência de neutralidade. O segundo é que a lei assegura a revisão, mas não exige que seja feita por pessoa natural. O terceiro é que o direito depende de iniciativa do trabalhador, que, subordinado e sem conhecer as métricas, dificilmente a exerce; a auditoria, por sua vez, depende da recusa do controlador e da atuação da ANPD. O art. 20 apenas permite ao trabalhador pedir explicações, e não afirma a premissa de que quem detém o poder e a informação deve demonstrar a legitimidade de suas decisões.

4.2 A norma regulamentadora nº 1 e o gerenciamento de riscos psicossociais

Como visto na seção 3.3, a NR-1 exige, desde 26 de maio de 2026, que o gerenciamento de riscos contemple os fatores psicossociais (BRASIL, 2024). Isso permite tratar a pressão por metas como risco a ser identificado, avaliado e controlado, e não como questão de gestão. A lacuna é que a norma não trata expressamente dos sistemas automatizados de metas e avaliação a organização pode cumpri-la formalmente sem inventariar os efeitos do algoritmo, como a recalibração contínua das metas ou a exposição permanente a rankings. A decisão do TRT da 2ª Região, que considerou ineficazes medidas preventivas meramente formais, indica que a efetividade importa (BRASIL, 2026, ROT-1000485-78.2025.5.02.0081, 1ª Turma). Falta norma que determine o que examinar quando o risco é produzido por um sistema.

4.3 O PL nº 2.338/2023 e a proteção em face da automação

O Projeto de Lei nº 2.338/2023, aprovado pelo Senado em 10 de dezembro de 2024, é a proposta mais relevante para o tema. Seu art. 14, III, classifica como de alto risco o sistema de IA empregado em recrutamento, promoções, cessação de relações contratuais, "avaliação do desempenho e do comportamento das pessoas afetadas nas áreas de emprego" e gestão de trabalhadores (BRASIL, 2023). A classificação confirma a hipótese deste capítulo: avaliar e gerir pessoas por IA é atividade de risco elevado. Ao aplicador, que na relação de trabalho é o empregador, o projeto atribui deveres relevantes, como a avaliação de impacto algorítmico (art. 25), o direito do afetado à explicação, à contestação e à revisão humana (art. 6º) e a inversão do ônus da prova em favor do hipossuficiente (art. 37). Os princípios da lei incluem a proteção do trabalho e do trabalhador (art. 3º, I).

Três limites, porém, persistem. O primeiro é que a lista do art. 14 depende de regulamentação, e o aplicador pode pedir à autoridade setorial que reconheça que seu sistema não é de alto risco (art. 16, § 2º). O segundo é que o parágrafo único exclui o sistema usado como "tecnologia intermediária que não influencie ou determine resultado ou decisão". Reproduz-se aqui a dificuldade do art. 20 da LGPD, pois o sistema que apenas "auxilia" o gestor pode escapar da classificação, ainda que molde sua decisão. O terceiro é que a proteção específica do trabalhador se restringe a diretrizes, voltadas ao deslocamento de emprego e à capacitação, e à valorização da negociação coletiva (art. 58), sem regra própria sobre metas ou avaliação. O projeto aguarda parecer do relator na Comissão Especial da Câmara.

Soma-se a esse cenário o julgamento da Ação Direta de Inconstitucionalidade por Omissão nº 73, em que o Supremo Tribunal Federal, por unanimidade, reconheceu, em 9 de outubro de 2025, a omissão inconstitucional do Congresso Nacional na regulamentação da proteção do trabalhador em face da automação, prevista no art. 7º, XXVII, da Constituição Federal, fixando o prazo de 24 meses para que o Poder Legislativo supra a omissão (BRASIL, 2025, ADO 73/DF). A decisão não trata de IA nem de gestão algorítmica, mas afirma o dever de legislar e ajuda a tratar o tema como dever de proteção.

5 CRITÉRIOS PARA O USO RESPONSÁVEL DA IA NA GESTÃO DO TRABALHO: LIMITES AO PODER DOS ALGORITMOS

As lacunas examinadas não deixam o trabalhador desprotegido, mas deslocam a pergunta: em vez de esperar por uma lei específica, cabe indagar quais critérios o princípio da proteção já permite extrair do ordenamento para orientar o uso da IA em metas e avaliação de desempenho. Os critérios a seguir derivam das três premissas da seção 2.1 e respondem à questão que orienta esta obra, pois indicam o que se deve exigir para que a IA seja usada de modo responsável nas relações de trabalho. Organizam-se em limites materiais, procedimentais e de responsabilização.

5.1 Limites materiais: o que o sistema não pode fazer

Os limites materiais decorrem da primeira e da segunda premissas e retomam os critérios da seção 3.1. O primeiro é a viabilidade da meta. A meta calculada automaticamente continua sendo exercício do poder diretivo e deve ser alcançável em condições normais de trabalho. Metas recalculadas a cada ciclo a partir do desempenho dos melhores, ou escaladas sem parâmetros humanamente sustentáveis, deixam de ser instrumento de organização e passam a ser mecanismo de exaustão. O risco do negócio, que a CLT atribui ao empregador (BRASIL, 1943, art. 2º), não pode ser transferido ao trabalhador por metas que variam conforme o desempenho do próprio grupo.

O segundo é a forma da cobrança. A exposição contínua e em tempo real de rankings, o alerta permanente e a comparação ostensiva podem impor pressão incompatível com a dignidade, ainda que a finalidade seja legítima. Como indica a divergência jurisprudencial da seção 3.2, o ranking não é ilícito em si; importam a intensidade da exposição, o contexto e a presença de medidas de prevenção.

O terceiro é o efeito sobre a saúde: o sistema de metas e avaliação deve ser tratado como fonte de risco psicossocial, identificada e controlada no gerenciamento de riscos ocupacionais (BRASIL, 2024). O quarto é a não discriminação: métricas aparentemente neutras, como disponibilidade, ritmo ou tempo de resposta, podem desfavorecer pessoas com deficiência, com responsabilidades de cuidado ou com condições de saúde, caracterizando discriminação indireta (BRASIL, 1995). O sistema deve ser testado para esse efeito antes de ser usado e monitorado depois.

5.2 Limites procedimentais: como o poder deve ser exercido

Os limites procedimentais decorrem da terceira premissa: se o empregador detém o sistema e as informações, cabe a ele demonstrar a legitimidade de suas decisões. Exige-se, primeiro, transparência: o trabalhador deve saber quando é avaliado por sistema automatizado, quais dados são coletados e quais critérios determinam suas metas e notas. O dever decorre da boa-fé objetiva (BRASIL, 2002, art. 422) e é reforçado pelo princípio da transparência da LGPD (BRASIL, 2018, art. 6º, VI). O segredo comercial do fornecedor não pode ser oposto ao trabalhador de modo a impedir a verificação da decisão.

Exige-se, segundo a revisão humana efetiva. Não basta que um gestor referende a decisão do sistema: ele deve ter competência, informação e autonomia para divergir. O critério dialoga com a lacuna do art. 20 da LGPD, restrito às decisões "unicamente" automatizadas, e com o parágrafo único do art. 14 do PL 2.338/2023, que exclui o sistema que não "influencie ou determine" a decisão. Defende-se aqui que o critério deve ser a influência real do sistema, e não a presença formal de uma pessoa na cadeia decisória; se a nota ou a meta orienta a promoção, a advertência ou o desligamento, a decisão deve ser tratada como automatizada para fins de revisão.

Exige-se, terceiro, a avaliação prévia e periódica de impacto, inspirada na avaliação de impacto algorítmico do projeto (BRASIL, 2023, art. 25), que examine os efeitos sobre a saúde, a igualdade e a razoabilidade das metas e seja refeita a cada mudança tecnológica ou organizacional.

Exige-se, por fim, a participação coletiva. A introdução e a alteração de sistemas de gestão por IA afetam o conjunto dos trabalhadores e devem ser objeto de informação e negociação com os sindicatos: a assimetria informacional é individual e coletiva, e a representação coletiva é meio relevante de compensá-la (MARQUES; MARTINEZ NETO, 2022). O PL 2.338/2023 apenas indica a valorização da negociação coletiva como diretriz (art. 58, III), o que reforça a importância de que as normas coletivas explorem o tema.

5.3 Responsabilização de quem desenvolve e de quem utiliza

Os critérios só têm eficácia se houver consequência para seu descumprimento. Na relação de emprego, o responsável perante o trabalhador é o empregador, que assume os riscos da atividade e responde pelos instrumentos que escolhe, inclusive os fornecidos por terceiros (BRASIL, 1943, arts. 2º e 6º, parágrafo único). A terceirização tecnológica da decisão não terceiriza a responsabilidade: o empregador não pode alegar que desconhecia o funcionamento do sistema ou que a decisão foi "do algoritmo". O abuso do poder diretivo configura ato ilícito (BRASIL, 2002, art. 187) e, em caso de falta grave do empregador, pode ensejar a rescisão indireta (BRASIL, 1943, art. 483).

Isso não exclui a responsabilidade de quem desenvolve o sistema. O PL 2.338/2023 distingue o desenvolvedor do aplicador, impõe a cada um deveres de governança e documentação (arts. 18 e 25) e remete a responsabilidade civil ao Código Civil e à legislação especial, considerando o grau de autonomia e de risco do sistema (art. 36). Perante o trabalhador, porém, o empregador deve responder integralmente, resguardado seu direito de regresso contra o fornecedor, pois o trabalhador não tem como identificar a origem do defeito nem acesso ao sistema para prová-lo. Se o legislador adotar a responsabilidade solidária, estará em coerência com o princípio da proteção.

Por fim, a distribuição do ônus da prova deve acompanhar a posse da informação. O projeto prevê a inversão quando a vítima for hipossuficiente ou quando o funcionamento do sistema tornar a prova excessivamente onerosa (art. 37). No processo do trabalho, quem contesta uma nota ou meta não tem acesso aos dados que a produziram, e quem os detém deve exibi-los: é a terceira premissa aplicada ao processo.

6 CONSIDERAÇÕES FINAIS

Este capítulo buscou responder como o princípio da proteção do trabalhador pode orientar os limites ao uso da inteligência artificial na definição de metas e na avaliação de desempenho. Sustentou-se que o princípio, lido como critério de limitação do poder e não apenas como regra de interpretação, impõe três exigências: a dignidade do trabalhador limita o poder diretivo, esse limite alcança os meios tecnológicos de exercê-lo, e quem detém o poder e a informação deve demonstrar a legitimidade de suas decisões.

A análise mostrou que o gerenciamento algorítmico altera o exercício do poder diretivo em quatro aspectos: a continuidade, a opacidade, a assimetria de informação e a aparência de neutralidade. Os riscos daí decorrentes, como metas inalcançáveis, rankings permanentes, discriminação indireta e adoecimento psíquico, têm correspondência em parâmetros já usados pela jurisprudência, embora não existam, até onde se pôde verificar, decisões trabalhistas brasileiras sobre metas definidas por algoritmos. As normas existentes, como o art. 20 da LGPD e a NR-1, oferecem proteção parcial, e o PL 2.338/2023, que classifica a avaliação de trabalhadores como de alto risco, ainda não foi aprovado. Diante disso, propôs-se um conjunto de critérios: limites materiais (viabilidade das metas, forma da cobrança, proteção da saúde e não discriminação) e procedimentais (transparência, revisão humana efetiva, avaliação de impacto e participação coletiva), além da responsabilização do empregador e do ônus da prova a quem detém a informação.

Quanto à questão que orienta esta obra, a IA transforma as relações sociais ao deslocar decisões antes tomadas entre pessoas para sistemas opacos e contínuos, o que é especialmente sensível no trabalho, onde há subordinação. Os critérios que devem orientar seu uso responsável são a centralidade da pessoa, a transparência, a possibilidade de contestação e de revisão humana, a prevenção de danos e a responsabilização de quem decide e de quem lucra com o sistema. Tais critérios não impedem a inovação. Fixam as condições para que ela seja compatível com a dignidade de quem trabalha.

Como limitações, o estudo se restringiu ao direito brasileiro e trabalhou com hipóteses sobre o funcionamento dos sistemas, que pesquisas empíricas futuras poderão confirmar ou corrigir.

REFERÊNCIAS BIBLIOGRÁFICAS

BARROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. **California Law Review**, Berkeley, v. 104, n. 3, p. 671-732, 2016.

BOWDLER, Mairi; LAHTI, Heidi; JELENKO, Marie; BURESTI, Giuliana; VALTONEN, Teppo. Algorithmic management and psychosocial risks at work: an emerging occupational safety and health challenge. **Scandinavian Journal of Work, Environment & Health**, v. 52, n. 1, p. 1-5, 2026. DOI: 10.5271/sjweh.4270.

BRASIL. **Decreto-Lei nº 5.452, de 1º de maio de 1943**. Aprova a Consolidação das Leis do Trabalho. Diário Oficial da União: Brasília, DF, 9 ago. 1943.

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: 8 nov. 2026.

BRASIL. **Lei nº 8.213, de 24 de julho de 1991**. Dispõe sobre os Planos de Benefícios da Previdência Social e dá outras providências. Diário Oficial da União: Brasília, DF, 25 jul. 1991.

BRASIL. **Lei nº 9.029, de 13 de abril de 1995**. Proíbe a exigência de atestados de gravidez e esterilização, e outras práticas discriminatórias, para efeitos admissionais ou de permanência da relação jurídica de trabalho, e dá outras providências. Diário Oficial da União: Brasília, DF, 17 abr. 1995.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Diário Oficial da União: Brasília, DF, 11 jan. 2002.

BRASIL. **Lei nº 12.551, de 15 de dezembro de 2011**. Altera o art. 6º da Consolidação das Leis do Trabalho (CLT). Diário Oficial da União: Brasília, DF, 16 dez. 2011.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: Brasília, DF, 15 ago. 2018.

BRASIL. Tribunal Superior do Trabalho (2. Turma). Recurso de Revista n. 528-74.2011.5.09.0001. Relator: Ministro Renato de Lacerda Paiva. Julgado em: 16 nov. 2015. Publicado no DEJT em: 27 nov. 2015.

BRASIL. Tribunal Superior do Trabalho (3. Turma). Recurso de Revista n. 1485-42.2010.5.09.0088. Relator: Ministro Mauricio Godinho Delgado. Julgado em: 13 set. 2017. Publicado no DEJT em: 15 set. 2017a.

BRASIL. Tribunal Superior do Trabalho (4. Turma). Recurso de Revista n. 68300-89.2009.5.09.0012. Relator: Ministro João Oreste Dalazen. Julgado em: 22 mar. 2017. Publicado no DEJT em: 25 ago. 2017b.

BRASIL. Congresso Nacional. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF:

Senado Federal, 2023. Texto remetido à Câmara dos Deputados em 17 mar. 2025. Disponível em: https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262. Acesso em: 8 nov. 2026.

BRASIL. Ministério do Trabalho e Emprego. **Portaria MTE nº 1.419, de 27 de agosto de 2024**. Altera a Norma Regulamentadora nº 1 (NR-1). Diário Oficial da União: Brasília, DF, 2024. Disponível em: https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?captchafield=firstAccess&data=28%2F08%2F2024&jornal=515&pagina=131. Acesso em: 8 nov. 2026.

BRASIL. Ministério do Trabalho e Emprego. **Portaria MTE nº 765, de 15 de maio de 2025**. Prorroga o prazo de início de vigência da nova redação do capítulo "1.5 Gerenciamento de riscos ocupacionais" da Norma Regulamentadora nº 1 (NR-1) – Disposições Gerais e Gerenciamento de Riscos Ocupacionais. *Diário Oficial da União*: seção 1, Brasília, DF, ano 163, n. 91, p. 87, 16 maio 2025. Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-mte-n-765-de-15-de-maio-de-2025-629790381. Acesso em: 8 nov. 2026.

BRASIL. Supremo Tribunal Federal (Plenário). Ação Direta de Inconstitucionalidade por Omissão n. 73/DF. Relator: Ministro Luís Roberto Barroso. Julgado em: 9 out. 2025. Disponível em: https://legis.senado.leg.br/norma/41598427. Acesso em: 8 nov. 2026. 2025b.

BRASIL. Tribunal Regional do Trabalho da 18ª Região (2. Turma). **Recurso Ordinário Trabalhista nº 0010553-89.2024.5.18.0018**. Relator: Desembargador Platon Teixeira de Azevedo Filho. Acórdão publicado em 17 out. 2025. *Informativo de Precedentes*, n. 280. Disponível em: https://bibliotecadigital.trt18.jus.br/xmlui/bitstream/handle/bdtrt18/34658/Informativo%20computador%20280.pdf?isAllowed=y&sequence=1. Acesso em: 8 nov. 2026.

BRASIL. Tribunal Regional do Trabalho da 2ª Região (1. Turma). **Recurso Ordinário Trabalhista nº 1000485-78.2025.5.02.0081**. Relator: Desembargador Willy Santilli. Acórdão assinado eletronicamente em 24 abr. 2026. Disponível em: https://arq.migalhas.com.br/arquivos/2026/5/B43558C337F6A1_Documento_d6f2c46-.pdf. Acesso em: 8 nov. 2026.

BRASIL. Tribunal Superior do Trabalho (6. Turma). **Agravo em Recurso de Revista com Agravo de Instrumento nº 1001166-22.2016.5.02.0030**. Relatora: Ministra Kátia Magalhães Arruda. Julgado em 4 fev. 2026. Brasília, DF: Tribunal Superior do Trabalho, 2026.

BRASIL. Tribunal Superior do Trabalho (8. Turma). Recurso de Revista com Agravo n. 12520-13.2016.5.15.0026. Relator: Ministro Evandro Pereira Valadão Lopes. 2026c.

CEFALIELLO, Aude; MOORE, Phoebe V.; DONOGHUE, Robert. Making algorithmic management safe and healthy for workers: addressing psychosocial risks in new legal provisions. *European Labour Law Journal*, v. 14, n. 2, p. 192-210, 2023. DOI: 10.1177/20319525231167476. Disponível em: https://doi.org/10.1177/20319525231167476. Acesso em: 8 nov. 2026.

DASTIN, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. **Reuters**, 10 out. 2018.

DELGADO, Mauricio Godinho. **Curso de direito do trabalho**. 18. ed. São Paulo: LTr, 2019.

KALIL, Renan Bernardi. **A regulação do trabalho via plataformas digitais**. São Paulo: Blucher, 2020. DOI: 10.5151/9786555500295.

KELLOGG, Katherine C.; VALENTINE, Melissa A.; CHRISTIN, Angèle. Algorithms at work: the new contested terrain of control. **Academy of Management Annals**, v. 14, n. 1, p. 366-410, 2020. DOI: 10.5465/annals.2018.0174.

MARQUES, Fabíola; MARTINEZ NETO, Aldo Augusto. Vieses algorítmicos, direitos fundamentais e os sindicatos. **Revista Jurídica Luso-Brasileira**, ano 8, n. 6, p. 707-729, 2022.

MIZIARA, Raphael. **Discriminação algorítmica e direito do trabalho**: condições e limites jurídicos para o uso da inteligência artificial nas relações de trabalho. 2024. 649 f. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade de São Paulo, São Paulo, 2024.

PINTO JUNIOR, Amaury Rodrigues; BOMFIM, Vólia. Riscos psicossociais e saúde mental no trabalho: impactos da atualização da NR-1. **Revista do Tribunal Superior do Trabalho**, Porto Alegre, v. 92, n. 1, p. 19-30, 2026.

PLÁ RODRIGUEZ, Américo. **Princípios de direito do trabalho**. 3. ed. São Paulo: LTr, 2000.

RODRIGUES, Priscila Lauande. A Convenção nº 193 da OIT sobre trabalho decente na economia de plataformas: primeiras reflexões sobre o novo marco normativo internacional. **Revista do Tribunal Superior do Trabalho**, v. 92, n. 2, p. 45, 2026. DOI: 10.70405/rtst.v92i2.303.

SANTOS, Rodrigo Coimbra; GRAMINHO, Vivian Maria Caxambu. Discriminação algorítmica nas relações de trabalho e princípios da LGPD. **Seqüência: Estudos Jurídicos e Políticos**, v. 45, n. 96, p. 1-31, 2024. DOI: 10.5007/2177-7055.2024.e96294.

TEIXEIRA, Sergio Torres; PIMENTEL, Alexandre Freire; PUGLIESI, Camila Crasto. Discriminação algorítmica nas relações de trabalho. **Revista do Tribunal Superior do Trabalho**, v. 91, n. 2, p. 116-142, 2025. DOI: 10.70405/rtst.v91i2.163.

TTL (Finnish Institute of Occupational Health); EU-OSHA. **Algorithmic management is becoming more common in the Nordics**: researchers warn about its impact on workers' well-being. Press release, 8 set. 2025. Disponível em: https://www.ttl.fi/en/topical/press-release/report-algorithmic-management-is-becoming-more-common-in-the-nordics-researchers-warn-about-its. Acesso em: 8 nov. 2026.

Inteligência Artificial e Jurisdição: Limites Constitucionais ao Uso de Sistemas Automatizados no Poder Judiciário Brasileiro

Artificial Intelligence and Jurisdiction: Constitutional Limits on the Use of Automated Systems in the Brazilian Judiciary

*Dante Batista Silva**

RESUMO

A incorporação da inteligência artificial ao Poder Judiciário brasileiro tem sido apresentada como resposta à sobrecarga processual e à demanda social por celeridade. Este artigo examina em que medida o emprego de sistemas automatizados na atividade jurisdicional é compatível com os direitos fundamentais e com as garantias constitucionais do processo. Parte-se da compreensão da jurisdição como poder, função e atividade para identificar como a tecnologia repercute em cada uma dessas dimensões. A pesquisa adota o método dedutivo, com abordagem qualitativa e técnicas de pesquisa bibliográfica e documental, a partir da doutrina processual e constitucional (Cintra; Grinover; Dinamarco, 1998; Russell; Norvig, 2013; Susskind, 2019) da legislação (Brasil, 2015), das resoluções do Conselho Nacional de Justiça (Brasil, 2020 e 2025) e de estudos sobre as ferramentas utilizadas pelo Supremo Tribunal Federal (Santos, 2024; Wotkoski, 2024), em especial o Projeto Victor. Os resultados indicam que a informatização dos tribunais brasileiros, tardia e orientada pela contenção do volume de processos, produziu ganhos de eficiência, mas que esses ganhos não podem ser dissociados da imparcialidade, da impessoalidade, da transparência e do devido processo legal. Conclui-se que a inteligência artificial é constitucionalmente admissível como instrumento de apoio à atividade jurisdicional, desde que submetida a supervisão humana efetiva, a mecanismos de revisão e a governança proporcional ao risco, como exige a Resolução CNJ n.º 615/2025, e que a pretensa neutralidade algorítmica deve dar lugar ao controle das escolhas valorativas que orientam a construção desses sistemas.

Palavras-chave: Inteligência Artificial; Jurisdição; Direitos Fundamentais; Devido Processo Legal; Projeto Victor.

* Professor do curso de Direito da Faculdade Lusófona de São Paulo (FLSP), com 12 anos de experiência docente na graduação e na pós-graduação. Membro do Grupo de Pesquisa em História da Educação e Infância (NIPHEI - UERJ). Doutor em Educação pela Universidade do Estado do Rio de Janeiro (UERJ). Mestre em Educação pela Universidade São Francisco (USF). Especialista em Direito Empresarial, com ênfase em Questões Tributárias, pela Universidade Metodista de Piracicaba (UNIMEP). Bacharel em Direito pela Universidade Bandeirante de São Paulo (UNIBAN).

ABSTRACT

The incorporation of artificial intelligence into the Brazilian Judiciary has been presented as a response to case overload and to the social demand for speedy proceedings. This article examines the extent to which the use of automated systems in judicial activity is compatible with fundamental rights and with the constitutional guarantees of due process. It starts from the understanding of jurisdiction as power, function and activity in order to identify how technology affects each of these dimensions. The research adopts the deductive method, with a qualitative approach and bibliographic and documentary research techniques, drawing on procedural and constitutional doctrine, legislation, resolutions of the National Council of Justice and studies on the tools used by the Federal Supreme Court, especially the Victor Project. The results indicate that the computerization of Brazilian courts, which was late and driven by the need to contain the volume of cases, has produced efficiency gains, but that these gains cannot be dissociated from impartiality, impersonality, transparency and due process of law. It concludes that artificial intelligence is constitutionally admissible as a tool to support judicial activity, provided that it is subject to effective human oversight, review mechanisms and risk-proportionate governance, as required by CNJ Resolution No. 615/2025, and that the alleged algorithmic neutrality must give way to the control of the value-laden choices that guide the design of these systems.

Keywords: Artificial Intelligence; Jurisdiction; Fundamental Rights; Due Process of Law; Victor Project.

1 INTRODUÇÃO

Poucas inovações alteraram tão rapidamente a rotina dos tribunais quanto a inteligência artificial (IA). Sistemas capazes de classificar recursos, agrupar processos semelhantes e sugerir minutas passaram a integrar o cotidiano do Poder Judiciário brasileiro, impulsionados por um diagnóstico antigo: o volume de demandas supera a capacidade de resposta das estruturas judiciárias, e a morosidade permanece como a crítica mais recorrente dos jurisdicionados.

A legislação processual tentou enfrentar esse quadro. O Código de Processo Civil (Brasil, 2015) trouxe inovações que alimentaram a expectativa de reduzir o volume de demandas e tornar o Judiciário mais eficiente. Passados alguns anos de vigência, no entanto, essa promessa não se concretizou. Nem o estímulo aos métodos consensuais de solução de conflitos nem a transferência de certas atribuições às serventias extrajudiciais bastaram para afastar a percepção de que a prestação jurisdicional continua tardia. É nesse contexto que a inteligência artificial passa a ser vista como solução em potencial.

A tecnologia, contudo, já não se restringe a tarefas administrativas. A IA passa a participar da tomada de decisão judicial, desde despachos e decisões interlocutórias até sentenças e acórdãos. Diante disso, torna-se

imprescindível delimitar com nitidez as balizas constitucionais que devem conter a atuação do Estado-juiz, sobretudo porque o próprio conceito de jurisdição atravessa um processo de ressignificação (Siqueira; Morais; Santos, 2022, p. 10-11).

Assim, o presente artigo propõe-se a responder ao seguinte problema: em que medida o uso da inteligência artificial na atividade jurisdicional é compatível com os direitos fundamentais e com as garantias constitucionais do processo? Trabalha-se com a hipótese de que a IA é admissível como instrumento de apoio ao julgador, mas não como substituta do juízo humano, devendo seu emprego subordinar-se à supervisão humana, à transparência e à possibilidade de revisão.

O objetivo geral consiste em analisar os limites constitucionais ao uso da inteligência artificial no Poder Judiciário. Como objetivos específicos, busca-se: a) apresentar os fundamentos conceituais da inteligência artificial e sua definição normativa no âmbito do Judiciário; b) examinar as repercussões da tecnologia sobre as dimensões da jurisdição; c) descrever a trajetória de informatização dos tribunais brasileiros e as ferramentas adotadas pelo Supremo Tribunal Federal (STF); e d) identificar os princípios constitucionais e os parâmetros regulatórios que condicionam o emprego desses sistemas.

A relevância do tema decorre da expansão acelerada dessas ferramentas. Levantamento do Conselho Nacional de Justiça (CNJ) registrou 178 iniciativas de IA em tribunais e conselhos, e quase metade dos órgãos participantes declarou utilizar ferramentas generativas (Conselho Nacional de Justiça, 2025a). Uma transformação dessa amplitude, que alcança o núcleo da função de julgar, exige exame jurídico cuidadoso.

2 PERCURSO METODOLÓGICO

A pesquisa caracteriza-se, quanto à natureza, como básica e, quanto aos objetivos, como exploratória e descritiva, uma vez que busca compreender um fenômeno recente e ainda em consolidação no ordenamento jurídico brasileiro. Adota-se abordagem qualitativa, pois o interesse recai sobre a interpretação de normas, conceitos e argumentos doutrinários, e não sobre a mensuração estatística do fenômeno.

Quanto ao método de abordagem, emprega-se o dedutivo: parte-se de premissas gerais, extraídas da teoria da jurisdição e dos princípios constitucionais, para avaliar o caso particular do uso da inteligência artificial pelos tribunais. Como procedimentos técnicos, utilizam-se a pesquisa bibliográfica e a pesquisa documental.

A pesquisa bibliográfica abrangeu obras de referência em inteligência artificial (Russell; Norvig, 2013; Susskind, 2019) e em teoria geral do processo (Cintra; Grinover; Dinamarco, 1998), artigos publicados em periódicos jurídicos e dissertações recentes sobre o uso da IA no Supremo Tribunal Federal (Santos, 2024; Wotkoski, 2024). A pesquisa documental compreendeu a Constituição Federal de 1988 (Brasil, 1988), o Código de Processo Civil (Brasil 2015), a legislação sobre a informatização do processo judicial, as Resoluções n.º 332/2020 e n.º 615/2025 do Conselho Nacional de Justiça, relatórios institucionais e notícias oficiais dos tribunais.

Na seleção das fontes sobre o uso da IA no Judiciário, priorizou-se a produção posterior à Resolução CNJ n.º 332/2020, primeira norma do Conselho dedicada ao tema. A análise do material deu-se por leitura crítica e pelo confronto entre posições doutrinárias, de modo a evidenciar convergências e divergências, como as que envolvem a suposta neutralidade dos algoritmos. Os dados quantitativos mencionados ao longo do texto, a exemplo dos tempos de decisão associados ao Projeto Victor, provêm de fontes secundárias e são empregados como elementos de ilustração do argumento, e não como resultado de levantamento empírico próprio.

3 INTELIGÊNCIA ARTIFICIAL: FUNDAMENTOS CONCEITUAIS

Antes de examinar o uso da inteligência artificial pelo Judiciário, convém delimitar o que se entende por essa tecnologia. A tarefa não é simples, porque a expressão abriga tradições distintas na ciência da computação e, mais recentemente, ganhou contornos normativos próprios.

3.1 Origens e conceito de inteligência artificial

Russell e Norvig situam a inteligência artificial no prolongamento de uma indagação muito antiga, a de compreender como o ser humano pensa, e destacam a singularidade do campo em relação às demais ciências:

> Denominamos nossa espécie Homo sapiens — homem sábio — porque nossa inteligência é tão importante para nós. Durante milhares de anos, procuramos entender como pensamos, isto é, como um mero punhado de matéria pode perceber, compreender, prever e manipular um mundo muito maior e mais complicado que ela própria. O campo da inteligência artificial, ou IA, vai ainda mais além: ele tenta não

> apenas compreender, mas também construir entidades inteligentes. A IA é um dos campos mais recentes em ciências e engenharia. O trabalho começou logo após a Segunda Guerra Mundial, e o próprio nome foi cunhado em 1956. Juntamente com a biologia molecular, a IA é citada regularmente como "o campo em que eu mais gostaria de estar" por cientistas de outras disciplinas. Um aluno de física pode argumentar, com boa dose de razão, que todas as boas ideias já foram desenvolvidas por Galileu, Newton, Einstein e o resto. IA, por outro lado, ainda tem espaço para vários Einsteins e Edisons em tempo integral (Russell; Norvig, 2013, p. 24).

Destacar o interesse que a inteligência artificial desperta não equivale, contudo, a defini-la. Ao organizar as conceituações disponíveis na literatura, Russell e Norvig (2013, p. 24) as distribuem em dois eixos: o primeiro privilegia os processos de pensamento e raciocínio; o segundo volta-se para o comportamento.

A preocupação com o raciocínio correto, aliás, antecede em muitos séculos a computação. Aristóteles figura entre os primeiros a buscar sistematizar o que seria um raciocínio irrefutável. Seus silogismos funcionavam como esquemas argumentativos nos quais premissas verdadeiras conduzem necessariamente a uma conclusão verdadeira. É o caso da conhecida inferência segundo a qual, sendo Sócrates homem e todo homem mortal, também Sócrates há de ser mortal. A intenção era que essas leis do pensamento regessem o próprio funcionamento da mente, e foi do estudo delas que surgiu a lógica (Russell; Norvig, 2013, p. 27-28).

A abordagem que mais interessa ao presente estudo, porém, é a do agente racional. Em sua acepção mais elementar, agente é aquilo que age; o termo provém do latim agere, "fazer". Todo programa de computador executa alguma tarefa, mas do agente computacional espera-se mais: que atue com autonomia, perceba o ambiente em que está inserido, mantenha-se em funcionamento por períodos prolongados, ajuste-se a mudanças e seja capaz de estabelecer e perseguir objetivos. Nessa perspectiva, racional é o agente que busca o melhor resultado possível ou, sob condições de incerteza, o melhor resultado esperado (Russell; Norvig, 2013). É essa noção, centrada na ação e no resultado, que melhor descreve os sistemas hoje empregados pelos tribunais, concebidos para executar tarefas com algum grau de autonomia.

3.2 A inteligência artificial no contexto jurídico e sua definição normativa

Transposta para o campo do Direito, essa compreensão ganha contornos mais práticos. Susskind (2019, p. 265) propõe uma concepção ampla de inteligência artificial e aponta, como seu traço mais relevante, a substituição progressiva de tarefas humanas por máquinas. Essa leitura corresponde, em boa medida, à forma como advogados e analistas costumam se referir à IA no meio jurídico, isto é, como sistemas capazes de executar funções que antes dependiam do trabalho reflexivo de profissionais humanos. Nessa perspectiva, a IA compreende tanto capacidades análogas às humanas quanto habilidades que ultrapassam o desempenho humano (Susskind, 2019, p. 265).

No plano normativo, o Conselho Nacional de Justiça adotou definição igualmente abrangente. Nos termos da Resolução n.º 615/2025, considera-se sistema de inteligência artificial aquele que, operando a partir de uma máquina e com níveis variáveis de autonomia, processa dados ou informações recebidas para atingir finalidades explícitas ou implícitas, gerando resultados prováveis e coerentes, sob a forma de decisões, recomendações ou conteúdos aptos a produzir efeitos no ambiente virtual ou físico (Conselho Nacional de Justiça, 2025b).

A amplitude da definição é deliberada: ela alcança tanto ferramentas de classificação e triagem quanto sistemas generativos, capazes de produzir textos. É justamente por isso que se torna necessário examinar como a tecnologia se relaciona com a função jurisdicional, cujo exercício a Constituição reserva ao Estado-juiz.

4 JURISDIÇÃO E INTELIGÊNCIA ARTIFICIAL

Compreender os efeitos da inteligência artificial sobre a atividade judicial exige retornar ao conceito de jurisdição. Somente a partir dele é possível distinguir o que pode ser automatizado daquilo que integra o núcleo indelegável do ato de julgar.

4.1 A jurisdição como poder, função e atividade

A incorporação da inteligência artificial ao exercício da jurisdição repercute, de maneira distinta, em cada uma de suas dimensões, entendida a jurisdição, a um só tempo, como poder, função e atividade, que só se

manifestam legitimamente por meio do devido processo legal (Cintra; Grinover; Dinamarco, 1998, p. 89). Tomada como poder, a jurisdição exprime a própria soberania estatal, e o juiz a exerce não em nome próprio, mas como agente do Estado, razão pela qual não lhe é dado transferir a outrem as atribuições que a Constituição e a lei lhe confiaram (Cintra; Grinover; Dinamarco, 1998, p. 94).

Daí a pergunta inevitável: até que ponto seria legítimo atribuir parcelas dessa autoridade a sistemas automatizados? Sob o prisma da função, cuja finalidade é a pacificação dos conflitos, o cenário parece mais favorável, pois a tecnologia pode contribuir para abreviar a duração dos processos e conferir maior uniformidade às decisões e aos procedimentos. Já na dimensão da atividade, compreendida como o conjunto de atos praticados pelo juiz no processo, os desafios recaem sobre a transparência e a responsabilização. Ferramentas destinadas à análise de precedentes, ou mesmo à elaboração parcial de decisões e sentenças, tendem a tornar o trabalho jurisdicional mais eficiente, mas esse ganho não dispensa o escrutínio sobre quem responde pelo resultado e como ele foi alcançado.

4.2 Os pronunciamentos judiciais e os graus de automação

A distinção entre essas dimensões ganha concretude quando se observam os atos praticados pelo juiz. No ordenamento brasileiro, o art. 203 do Código de Processo Civil (Brasil, 2015) classifica os pronunciamentos do juiz em sentenças, decisões interlocutórias e despachos. Cada espécie desempenha papel próprio no processo e, justamente por isso, admite em graus distintos o emprego da inteligência artificial. Os despachos, desprovidos de conteúdo decisório, prestam-se com maior facilidade à automação; as decisões interlocutórias e as sentenças, que resolvem questões e exigem fundamentação, situam-se em zona de maior sensibilidade, pois nelas se concentra o exercício do poder jurisdicional.

Daí decorre que as ferramentas de IA precisam ser compatíveis com os direitos fundamentais. Elas também devem dispor de instrumentos de correção, supervisão e controle, sem os quais dificilmente conquistarão a confiança dos jurisdicionados. São esses parâmetros constitucionais que demarcam os limites a serem respeitados pelos profissionais do Direito ao recorrer a tais tecnologias (Siqueira; Morais; Santos, 2022, p. 3). Antes de examiná-los, porém, é preciso conhecer o caminho pelo qual a tecnologia chegou aos tribunais brasileiros.

5 A INFORMATIZAÇÃO DO PODER JUDICIÁRIO BRASILEIRO

A presença atual da inteligência artificial nos tribunais é o ponto mais recente de um processo de informatização iniciado décadas atrás. Reconstruir essa trajetória ajuda a compreender por que a tecnologia foi recebida, desde o início, sobretudo como instrumento de contenção do volume de processos.

5.1 Uma informatização tardia e orientada pelo volume de demandas

A chegada da tecnologia aos tribunais brasileiros guarda uma singularidade. Ela não resultou de um movimento gradual de modernização: veio tarde e já carregada de uma expectativa precisa, a de dar vazão ao trabalho acumulado. Santos (2024, p. 51) resume essa particularidade ao observar que "a informatização dos tribunais brasileiros é relativamente tardia e já pensada como uma solução para resolver o grande volume de trabalho do Poder Judiciário".

O atraso, segundo o autor, tem raízes estruturais. A formação de mão de obra especializada caminhou devagar: o primeiro curso superior de ciência da computação surgiu em 1968, e somente em 1996 o país ultrapassou a marca de cem cursos. Pesou também a Política Nacional de Informática, instituída pela Lei n.º 7.232/1984 (Brasil, 1984), que restringiu por oito anos a importação de computadores. Quando, enfim, formou-se uma base mínima, a difusão foi rápida. Os primeiros sistemas de gestão de informação surgiram nos tribunais ainda nos anos 1980; na década seguinte, os computadores passaram a integrar a rotina dos magistrados; e, em 1996, o Superior Tribunal de Justiça lançou seu portal eletrônico, com consulta à jurisprudência e ao andamento processual (Santos, 2024, p. 51).

A virada do século aprofundou esse movimento. A Lei n.º 10.259/2001 (Brasil, 2001), ao disciplinar os Juizados Especiais Federais, abriu caminho para o processo judicial eletrônico, em sintonia com a celeridade que orienta esse microssistema (Gomes; Nunes, 2022, p. 95). No mesmo ano, o Supremo Tribunal Federal firmou convênio com o Banco Central do Brasil para desenvolver um mecanismo de localização e bloqueio de bens, voltado a assegurar a efetividade das decisões judiciais. Surgia o BACENJUD, cujo nome combina "Banco Central" e "Judiciário", hoje substituído pelo SISBAJUD, o Sistema de Busca de Ativos do Poder Judiciário (Pessoa; Guimarães, 2022, p. 142).

Se a tecnologia demorou a chegar, sua expansão a partir da Reforma do Judiciário foi marcada pela pressa, num ambiente de explosão de demandas e de cobrança crescente por eficiência. Santos descreve esse período nos seguintes termos:

> Esse contexto de choque de demanda e pressão por maior eficiência compõe o cenário para a informatização do Poder Judiciário. Nos anos seguintes à reforma, o que se vê é um ciclo muito curto entre a criação e a adoção de tecnologias nos tribunais brasileiros. Pouco tempo depois da Emenda Constitucional 45/04, é editada a Lei 11.419/06, que permite a edição de diários de justiça eletrônicos, bem como estabelece as bases para o processo eletrônico. Já no ano seguinte, o Tribunal de Justiça de São Paulo inaugura as primeiras varas exclusivamente digitais e, em 2009, o Supremo Tribunal Federal torna obrigatório o uso do seu portal eletrônico para algumas ações originárias. Esse processo de virtualização segue em ritmo acelerado até que, em 2021, o Conselho Nacional de Justiça proíbe a abertura de processos físicos (Santos, 2024, p. 52).

A passagem evidencia que, no Brasil, a tecnologia foi incorporada menos como projeto de modernização planejado e mais como resposta à crise de demanda. O intervalo de quinze anos entre a Lei n.º 11.419/2006 (Brasil, 2006) e a vedação de novos processos físicos dá a medida dessa aceleração. Nesse período, uma série de recursos tecnológicos tornou-se parte indissociável da atuação jurisdicional, como a penhora eletrônica, a perícia digital, as audiências por videoconferência e os leilões virtuais. Essas ferramentas ampliaram a transparência e a agilidade dos procedimentos e ofereceram ao Judiciário meios concretos de dar cumprimento ao princípio da celeridade também na fase de execução das decisões (Gomes; Nunes, 2022, p. 95).

O avanço tecnológico próprio da sociedade da informação em rede repercutiu, assim, de forma expressiva na distribuição da justiça e no campo jurídico como um todo. Foram essas transformações que prepararam o terreno para que a inteligência artificial ganhasse espaço no sistema judicial, e em nenhum outro tribunal esse movimento é tão visível quanto no Supremo Tribunal Federal.

5.2 O Supremo Tribunal Federal e as ferramentas de inteligência artificial

A mesma lógica de contenção do volume ajuda a explicar por que a inteligência artificial ganhou espaço justamente no STF, onde o excesso de processos é problema antigo. Os diagnósticos de sobrecarga remontam aos anos 1960, e diversas medidas foram adotadas desde então para reduzir o fluxo de recursos. A mais recente é a repercussão geral, criada pela Emenda Constitucional n.º 45/2004 (Brasil, 2004) e hoje operada com o apoio do Projeto Victor. Para Santos (2024, p. 58), esse requisito processual cumpre a função de "atuar como um filtro de relevância para as pautas do STF".

A lógica do instituto é que cada questão constitucional seja enfrentada uma única vez. Primeiro, formula-se um tema que expresse a controvérsia de alcance transcendente. Em seguida, o Tribunal delibera sobre sua relevância, isto é, sobre a existência de repercussão geral. Se a resposta for positiva, fixa a tese correspondente (Santos, 2024, p. 58).

A busca por celeridade atravessa todo esse percurso. Ela figurou entre as principais inquietações sociais que motivaram a Reforma do Judiciário e, apesar de avanços recentes nas justiças estadual e federal, a demora no julgamento continua a ser a crítica mais recorrente dos jurisdicionados. Nesse contexto, Santos (2024, p. 90) considera natural que a inteligência artificial, "dotada da capacidade de reproduzir rapidamente o resultado de análises humanas", passe a ser vista como uma solução em potencial.

Entre as iniciativas de maior repercussão no país, sobressai o Projeto Victor, fruto da parceria entre o Supremo Tribunal Federal e a Universidade de Brasília (UnB). O sistema emprega inteligência artificial para identificar, nos recursos que chegam à Corte, os temas de repercussão geral a que se vinculam. Essa etapa de classificação é decisiva para a gestão do instituto e para o andamento dos recursos no Tribunal. A proposta central da ferramenta é reduzir o tempo de análise dos processos, tarefa que exige considerável sofisticação em aprendizado de máquina, o chamado machine learning (Gomes; Nunes, 2022, p. 95).

A segunda ferramenta de inteligência artificial adotada pelo STF é a RAFA 2030 (Redes Artificiais Focadas na Agenda 2030), lançada em maio de 2022. A solução classifica os processos conforme os Objetivos de Desenvolvimento Sustentável (ODS) previstos na Agenda 2030 da Organização das Nações Unidas (ONU) e, com isso, padroniza e cataloga o acervo à luz dos 17 objetivos que compõem a agenda (Supremo Tribunal Federal, 2022).

5.3 Os efeitos do Projeto Victor sobre a celeridade processual

Resta saber se a promessa de celeridade se confirma na prática. Santos (2024, p. 93) adverte que os números, isoladamente, não autorizam concluir que o Victor seja determinante para a redução do tempo de decisão, pois a ferramenta trabalha com temas muito frequentes em recursos padronizados, cenário em que os próprios servidores tendem a decidir mais rapidamente com o passar do tempo.

Para testar essa hipótese, o autor examinou as decisões proferidas entre 2018 e agosto de 2021, período anterior ao uso do Victor. Naquele intervalo, o tempo médio foi de 22,7 dias para decisões com temas não abrangidos pela ferramenta, de 18 dias para as que aplicavam apenas temas depois incorporados a ela e de 18,6 dias para as que combinavam as duas situações. O próprio autor reconhece um limite metodológico: "analisar períodos diferentes atrai a influência de outros fatores causais" (Santos, 2024, p. 93). Entre esses fatores estão o aperfeiçoamento das equipes de triagem, a adoção de outras ferramentas, como o Projeto Admissibilidade, e a queda expressiva no número de recursos remetidos ao STF, de 80.747 em 2018 para 49.622 em 2022.

Feitas essas ressalvas, os dados apontam que os processos com contribuição do Victor tiveram redução mais acentuada no tempo de decisão, de 18,6 para 11,1 dias, ao passo que nos demais a queda foi de 22,7 para 20,3 dias (Santos, 2024, p. 93). O resultado é favorável à ferramenta, mas não permite isolar seu efeito com segurança.

Mesmo com indícios positivos, o alcance do Victor permanece restrito. Santos (2024, p. 99) sustenta que seu uso amplo no Tribunal poderia convertê-lo em referência comum para as decisões sobre repercussão geral, além de permitir reunir avaliações de equipes distintas sobre a qualidade das sugestões, o que alimentaria o aprimoramento do algoritmo. Como não foi implantado nos gabinetes, o sistema "deixa de contribuir para o combate de divergências entre os gabinetes", bem como entre estes e a Presidência (Santos, 2024, p. 99).

5.4 Panorama do uso da inteligência artificial nos tribunais brasileiros

A experiência do STF não é isolada. Ao comentar a presença da inteligência artificial nos tribunais, o ministro Emmanoel Pereira destacou que "percebe-se aumento na eficiência do trabalho dos gabinetes dos ministros, o que contribui para a celeridade dos julgamentos dos processos a partir da técnica de agrupamento". Em direção semelhante, mas com ênfase

nos limites da ferramenta, Pedro Felipe de Oliveira Santos, ex-secretário-geral da Presidência do STF e vice-diretor da Escola de Magistratura Federal da 6ª Região, afirmou, a respeito do Victor, que "a inteligência artificial não substitui os servidores e os ministros nas análises dos recursos e na tomada de decisões, estando sempre sob a supervisão da inteligência humana" (Tribunal Superior do Trabalho, 2022).

Essas manifestações refletem um movimento mais amplo. Segundo Siqueira, Morais e Santos (2022, p. 3), pouco mais da metade dos tribunais brasileiros já dispunha, à época, de projetos de inteligência artificial em operação ou em desenvolvimento. O propósito dessas iniciativas é aperfeiçoar o funcionamento dos sistemas de processo eletrônico adotados pelo Judiciário, como Projudi, PJe, e-SAJ, E-US, e-JUR, JEF Virtual, PJD e JURIS e-Proc, a fim de tornar mais eficiente a gestão de processos e equipes e oferecer apoio aos magistrados na tomada de decisão.

Os dados mais recentes confirmam a expansão. A Pesquisa IA no Poder Judiciário 2024, divulgada pelo CNJ em 2025, contou com a participação de 92 dos 95 tribunais e conselhos, o que corresponde a 96,8% de adesão. O levantamento identificou 178 iniciativas, das quais 98 foram registradas pela primeira vez no ciclo de 2024, e 45,8% dos órgãos participantes declararam utilizar ferramentas de IA generativa, sobretudo em atividades ligadas à produção e ao tratamento de textos (Conselho Nacional de Justiça, 2025a). Esses números exigem leitura atenta: as 178 iniciativas não foram todas criadas em 2024, e o percentual de 45,8% diz respeito apenas aos órgãos que responderam à pesquisa.

Do ponto de vista institucional, a inteligência artificial tem sido apresentada como propulsora da transformação digital do Poder Judiciário, com o objetivo de tornar a Justiça um serviço mais acessível e eficiente. Esse movimento procura aproximar o Judiciário das demandas da população e ampliar o acesso à justiça, ao mesmo tempo em que busca dar maior celeridade à atividade jurisdicional e reduzir os custos operacionais desse serviço público (Conselho Nacional de Justiça, 2024, p. 218).

É inegável que o Processo Judicial Eletrônico, a inteligência artificial e outras ferramentas adotadas pelos tribunais se mostraram, num primeiro momento, instrumentos eficientes. Elas facilitaram a elaboração de relatórios e a organização de dados e ampliaram o acesso dos cidadãos às informações do Judiciário, com reflexos positivos sobre a transparência e a agilidade da prestação jurisdicional. O principal atrativo da tecnologia reside, portanto, na possibilidade de enfrentar o volume excessivo de

processos, reduzir o trabalho repetitivo dos operadores do Direito e ganhar eficiência. Seu uso nos tribunais já é uma realidade e, ao que tudo indica, irreversível.

A automação, contudo, não pode se converter em perda de qualidade da justiça, tampouco servir de pretexto para o desrespeito aos direitos fundamentais das partes. A função jurisdicional passou a contar com o auxílio de algoritmos, mas isso não significa que a tecnologia opere à margem do ordenamento: toda atividade desempenhada com o apoio da inteligência artificial deve encontrar respaldo nos princípios constitucionais. É desses limites que trata a seção seguinte.

6 LIMITES CONSTITUCIONAIS AO USO DA INTELIGÊNCIA ARTIFICIAL NA JURISDIÇÃO

Os ganhos de eficiência descritos na seção anterior só são legítimos se compatíveis com a ordem constitucional. Nesta seção, examinam-se os direitos fundamentais e as garantias processuais que condicionam o uso da IA, os princípios mais diretamente afetados pela tecnologia e a resposta regulatória do Conselho Nacional de Justiça.

6.1 Direitos fundamentais e garantias constitucionais do processo

Os princípios constitucionais orientam todo o ordenamento jurídico brasileiro e vinculam juízes, advogados e demais operadores do Direito. O mesmo vale para o emprego da inteligência artificial no Judiciário: a tecnologia representa inovação relevante nos tribunais, mas seu uso deve ser conduzido e delimitado por esses princípios.

A centralidade dos direitos fundamentais é fruto de uma trajetória histórica. Desde a independência em relação a Portugal, em 1822, o Brasil teve sete Constituições: as de 1824, 1891, 1934, 1937, 1946, 1967/1969 e 1988. Ao longo desses períodos, foram muitas as transformações institucionais, mas, segundo Ramos (2003, p. 28), foi a Constituição de 1988 que consolidou o compromisso com a garantia dos direitos fundamentais.

Esse compromisso foi reforçado pela proteção conferida aos direitos e garantias individuais como cláusula pétrea, nos termos do art. 60, § 4º, inciso IV, da Constituição Federal (Brasil, 1988), dispositivo que impede o Congresso Nacional de deliberar sobre proposta de emenda tendente a aboli-los. Nesse contexto, a atuação do Estado-juiz deve estar em sintonia com as garantias constitucionais dos litigantes e com

os valores objetivos consagrados pela Constituição. Por consequência, os projetos de inteligência artificial implantados no Judiciário também estão obrigados, desde a concepção até a execução, a assegurar a efetividade dos direitos fundamentais de natureza processual e o respeito às garantias constitucionais do processo (Passos, 2022, p. 98).

O uso da IA, portanto, não pode comprometer a imparcialidade, a legalidade e os direitos fundamentais, que sustentam a justiça em um regime democrático. O diálogo entre inovação tecnológica e garantias constitucionais é indispensável para que o processo judicial permaneça justo, transparente e socialmente legítimo.

6.2 Eficiência e razoável duração do processo.

Entre os princípios relacionados ao tema, o da eficiência, previsto no art. 37, caput, da Constituição Federal (Brasil, 1988), ocupa lugar de destaque, por ser o fundamento invocado com mais frequência para justificar a adoção da IA. Como integrante da Administração Pública, o Poder Judiciário está vinculado aos princípios da eficiência e da economicidade. Na prática, porém, convive com problemas estruturais, entre eles a distribuição desigual das equipes e a pouca transparência na gestão dos processos. Uma alocação mais racional da força de trabalho, aliada a boas práticas de gestão, poderia elevar o desempenho das unidades judiciárias e reduzir as disparidades entre magistrados, o que se refletiria em tratamento mais isonômico aos jurisdicionados (Wotkoski, 2024, p. 71).

A eficiência guarda relação direta com a razoável duração do processo e com a equidade no tratamento das partes. O desafio está em conciliar a autonomia dos magistrados com a exigência de uma gestão administrativa mais transparente. Para Wotkoski (2024, p. 71), a adoção de assessorias centralizadas e de métodos de organização mais apurados poderia abreviar o tempo de tramitação e tornar a gestão, ao mesmo tempo, mais eficiente e mais justa.

O princípio da razoável duração do processo, inscrito no art. 5º, inciso LXXVIII, da Constituição Federal (Brasil, 1988), garante às partes uma resposta judicial em prazo adequado, sem que o feito se arraste indevidamente. A afinidade com o princípio da eficiência é evidente: ambos buscam racionalizar o tempo e os atos processuais, de modo que os interesses em disputa sejam atendidos com justiça e sem demoras desnecessárias (Wotkoski, 2024, p. 72).

É nesse ponto que a inteligência artificial pode oferecer contribuição concreta. Ao assumir tarefas repetitivas e de baixa complexidade, os sistemas automatizados reduzem o tempo de tramitação e liberam juízes e servidores para as questões que demandam análise humana mais detida. Além disso, ao apontar etapas processuais dispensáveis, a IA pode tornar o fluxo de trabalho mais ágil e conter o acúmulo de processos, favorecendo tanto a eficiência quanto a duração razoável do processo (Wotkoski, 2024, p. 72).

6.3 Imparcialidade, impessoalidade e o problema da neutralidade algorítmica

Os ganhos de celeridade, contudo, não dispensam outras salvaguardas. O princípio da imparcialidade ocupa posição central no debate sobre o uso da inteligência artificial no Judiciário, já que as decisões automatizadas precisam preservar a equidade e a neutralidade esperadas de um sistema de justiça democrático. Para Wotkoski (2024, p. 75-76), isso exige que os algoritmos sejam concebidos e implementados a partir de critérios claros e objetivos, capazes de mitigar vieses e assegurar o respeito aos direitos fundamentais de todos os envolvidos.

A expectativa de neutralidade, porém, não é pacífica, e a experiência do Projeto Victor ajuda a compreender por quê. Segundo o relatório final do projeto, a equipe da UnB contou com seis pesquisadores da área jurídica e treze da área de tecnologia, além de membros do STF, sobre os quais há pouca informação, salvo a designação de dois servidores para acompanhar os trabalhos. No relatório elaborado por Luis Felipe Salomão e Caroline Tauk, a equipe foi considerada adequada quanto à diversidade e à interdisciplinaridade (Salomão; Tauk apud Santos, 2024, p. 99-104).

O ponto central, contudo, não está na qualificação da equipe, mas nas escolhas feitas ao longo do desenvolvimento. Santos não pretende julgar se tais escolhas foram acertadas; interessa-lhe demonstrar que poderiam ter sido outras, já que não existe uma única forma correta de construir uma aplicação de inteligência artificial, e cada alternativa altera o modo como os recursos serão avaliados. Daí sua conclusão de que a inteligência artificial não é neutra, pois sua construção envolve escolhas valorativas (Santos, 2024, p. 99-104).

O reconhecimento dessas escolhas é decisivo, porque ainda persiste uma percepção de neutralidade em torno da tecnologia. A literatura nacional divide-se a esse respeito. Parte dela admite a influência dos programadores e seus reflexos sobre a neutralidade e a imparcialidade dos sistemas; outra

parte sustenta que a IA favorece uma aplicação do direito livre de distorções e capaz de atenuar os vieses dos julgadores humanos (Santos, 2024, p. 99-104). A tensão entre essas leituras sugere que a imparcialidade não pode ser presumida a partir da natureza técnica da ferramenta. Ela precisa ser construída e verificada ao longo de todo o ciclo de desenvolvimento e uso, por meio de transparência sobre as escolhas realizadas, supervisão humana e possibilidade de revisão.

Na mesma direção, o princípio da impessoalidade mostra-se essencial à credibilidade do sistema judicial, sobretudo diante dos desafios trazidos pelo emprego da IA na tomada de decisões. A essa discussão soma-se o devido processo legal, cujo exame permite compreender de que modo a tecnologia pode repercutir sobre garantias processuais fundamentais. Daí a necessidade de que a adoção dessas ferramentas no Judiciário seja guiada por transparência, responsabilidade e equidade, de forma a proteger os direitos das partes e a integridade da justiça (Wotkoski, 2024, p. 79).

6.4 A regulação do CNJ: da Resolução n.º 332/2020 à Resolução n.º 615/2025

As preocupações expostas encontraram resposta normativa no Conselho Nacional de Justiça. A primeira norma do CNJ dedicada ao tema foi a Resolução n.º 332/2020, que estabeleceu parâmetros de ética, transparência e governança para o uso da inteligência artificial no Judiciário (Conselho Nacional de Justiça, 2020). A norma exigia que as soluções fossem transparentes e respeitassem a igualdade de tratamento, em consonância com o art. 5º da Constituição Federal, além de submeter seu uso a controle e a análise criteriosa.

Os riscos que motivaram essa preocupação continuam atuais. Um dos mais conhecidos é a possibilidade de que decisões automatizadas reproduzam preconceitos já presentes nos dados, com prejuízo à igualdade no processo. Outro é a opacidade: sem explicações acessíveis sobre o funcionamento dos algoritmos, torna-se mais difícil supervisioná-los e assegurar o direito de defesa, o que pode gerar violações constitucionais e afastar a atuação judicial dos princípios que fundamentam o ordenamento (Albergaria Neto; Lopes Junior, 2021).

A Resolução n.º 615/2025, que revogou a Resolução n.º 332/2020, aprofundou essa disciplina e partiu de uma compreensão qualitativa de eficiência, que não se resume à rapidez, mas envolve confiabilidade, transparência, segurança, possibilidade de revisão e preservação de direitos.

Nessa linha, a norma exige governança, monitoramento e supervisão humana proporcionais ao risco de cada solução de IA (Conselho Nacional de Justiça, 2025b). Essa cautela ganha peso nas aplicações generativas, cujos resultados devem permanecer sujeitos à validação humana e aos mecanismos institucionais de controle nela previstos.

A dimensão do acesso à justiça completa esse quadro. Aires (2023) chama atenção para a exclusão digital e lembra que a inovação tecnológica não pode ser avaliada apenas pelos ganhos internos de produtividade. Uma solução eficiente para o tribunal pode revelar-se inadequada se dificultar ao jurisdicionado compreender o processo, acessá-lo ou exercer seus direitos.

7 CONSIDERAÇÕES FINAIS

Este artigo propôs-se a investigar em que medida o uso da inteligência artificial na atividade jurisdicional é compatível com os direitos fundamentais e com as garantias constitucionais do processo. O percurso realizado permite confirmar a hipótese inicial: a IA é admissível como instrumento de apoio ao julgador, mas não como substituta do juízo humano, e sua legitimidade depende da submissão a supervisão humana efetiva, a transparência e a mecanismos de revisão.

A análise conceitual demonstrou que a inteligência artificial, compreendida como agente que atua com autonomia para alcançar resultados, recebeu do Conselho Nacional de Justiça definição deliberadamente ampla, apta a abranger tanto ferramentas de triagem quanto sistemas generativos. Essa amplitude impõe examinar a tecnologia à luz da jurisdição. Entendida como poder, função e atividade, a jurisdição revela tensões distintas: como poder, esbarra no princípio da indelegabilidade; como função, beneficia-se dos ganhos de celeridade e uniformidade; como atividade, exige transparência e responsabilização sobre os atos praticados com o auxílio de algoritmos. A classificação dos pronunciamentos judiciais no art. 203 do Código de Processo Civil reforça essa gradação, pois os despachos admitem automação com maior facilidade do que as decisões interlocutórias e as sentenças.

A reconstrução histórica evidenciou que a informatização do Judiciário brasileiro foi tardia e orientada, desde o início, pela contenção do volume de processos. O Projeto Victor sintetiza esse movimento: os dados disponíveis sugerem redução no tempo de decisão dos processos que contam com sua contribuição, mas não permitem isolar o efeito da ferramenta, e seu

alcance permanece restrito por não ter sido implantado nos gabinetes. A experiência mostra que a eficiência prometida pela tecnologia deve ser avaliada com cautela metodológica, e não presumida.

No plano dos limites constitucionais, verificou-se que a eficiência e a razoável duração do processo não podem ser perseguidas à custa da imparcialidade, da impessoalidade e do devido processo legal. O confronto entre as posições de Wotkoski (2024) e de Santos (2024) mostrou que a expectativa de neutralidade algorítmica é frágil, pois a construção dos sistemas envolve escolhas valorativas. Por isso, a imparcialidade deve ser construída e controlada ao longo de todo o ciclo de vida da ferramenta. A Resolução CNJ n.º 615/2025 caminha nessa direção ao exigir governança, monitoramento e supervisão humana proporcionais ao risco, embora sua efetividade dependa da implementação concreta pelos tribunais e da atenção ao acesso à justiça dos jurisdicionados digitalmente excluídos.

Reconhece-se, como limitação deste estudo, o caráter bibliográfico e documental da pesquisa, que se apoiou em dados secundários. Sugere-se, para investigações futuras, o exame empírico da aplicação da Resolução CNJ n.º 615/2025 nos tribunais, em especial quanto à transparência das ferramentas generativas e à efetividade da supervisão humana sobre as minutas produzidas com seu auxílio.

-

REFERÊNCIAS BIBLIOGRÁFICAS

AIRES, Andressa Soares Costa. Acesso à justiça, exclusão digital e a inteligência artificial no Poder Judiciário do Brasil: desafios e perspectivas. Revista do Tribunal Regional Federal da Primeira Região, [s. l.], v. 35, n. 1, p. 132-141, 2023. Disponível em: https://revista.trf1.jus.br/trf1/article/view/455. Acesso em: 2 nov. 2026.

ALBERGARIA NETO, Jason; LOPES JÚNIOR, Newton de Araújo. O princípio constitucional da razoável duração do processo e a implementação do processo judicial eletrônico como forma de garantir celeridade aos litígios. Revista Eletrônica Direito e Política, [s. l.], v. 16, n. 1, p. 129-156, 2021. DOI: 10.14210/rdp.v16n1.p129-156. Disponível em: https://periodicos.univali.br/index.php/rdp/article/view/17538. Acesso em: 2 nov. 2026.

BRASIL. Lei n.º 7.232, de 29 de outubro de 1984. Dispõe sobre a Política Nacional de Informática, e dá outras providências. Brasília, DF: Presidência da República, 1984.

BRASIL. Lei n.º 10.259, de 12 de julho de 2001. Dispõe sobre a instituição dos Juizados Especiais Cíveis e Criminais no âmbito da Justiça Federal. Brasília, DF: Presidência da República, 2001.

BRASIL. Emenda Constitucional n.º 45, de 30 de dezembro de 2004. Altera dispositivos dos arts. 5º, 36, 52, 92, 93, 95, 98, 99, 102, 103, 104, 105, 107, 109, 111, 112, 114, 115, 125, 126, 127, 128, 129, 134 e 168 da Constituição Federal, e acrescenta os arts. 103-A, 103-B, 111-A e 130-A, e dá outras providências. Brasília, DF: Presidência da República, 2004.

BRASIL. Lei n.º 11.419, de 19 de dezembro de 2006. Dispõe sobre a informatização do processo judicial; altera a Lei n.º 5.869, de 11 de janeiro de 1973 – Código de Processo Civil; e dá outras providências. Brasília, DF: Presidência da República, 2006.

BRASIL. Lei n.º 13.105, de 16 de março de 2015. Código de Processo Civil. Brasília, DF: Presidência da República, 2015.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidência da República, [2025]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 nov. 2026.

CINTRA, Antonio Carlos de Araújo; GRINOVER, Ada Pellegrini; DINAMARCO, Cândido Rangel. Teoria geral do processo. 14. ed. rev. e atual. São Paulo: Malheiros, 1998.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). Resolução n.º 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 e dá outras providências. Brasília, DF: CNJ, 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 2 nov. 2026.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). Justiça em números 2024. Brasília, DF: CNJ, 2024. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2024/05/justica-em-numeros-2024.pdf. Acesso em: 2 nov. 2026.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). Pesquisa aponta que uso de IA é tendência consolidada no Judiciário. Agência CNJ de Notícias, Brasília, DF, 16 set. 2025a. Disponível em: https://www.cnj.jus.br/pesquisa-aponta-que-uso-de-ia-e-tendencia-consolidada-no-judiciario/. Acesso em: 2 nov. 2026.

CONSELHO NACIONAL DE JUSTIÇA (Brasil). 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. Brasília, DF: CNJ, 2025b. Disponível em: https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf. Acesso em: 2 nov. 2026.

GOMES, Luísa Caroline; NUNES, Thâmylla da Cruz. Decisões judiciais conflitantes e o impacto da inteligência artificial na uniformização da jurisprudência. Revista de Direito e Atualidades, [s. l.], v. 1, n. 3, p. 86-104, jan. 2022. Disponível em: https://www.portaldeperiodicos.idp.edu.br/rda/article/view/6232. Acesso em: 28 out. 2026.

PASSOS, Hugo Assis. Inteligência artificial e a repercussão geral da questão constitucional: análise crítica e parâmetros de utilização. 2022. Tese (Doutorado em Direito Constitucional) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2022.

PESSOA, Flávia Moreira Guimarães; GUIMARÃES, Alessandro de Araújo. Novos paradigmas do acesso à justiça com o uso de inteligência artificial. In: ARAÚJO, Valter Shuenquener de; GOMES, Marcus Livio (coord.); CANEN, Doris (org.). Inteligência artificial e aplicabilidade prática no direito. [Brasília, DF]: Conselho Nacional de Justiça, 2022. p. 149-164. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2022/04/inteligencia-artificial-e-a-aplicabilidade-pratica-web-2022-03-11.pdf. Acesso em: 2 nov. 2026.

RAMOS, Edith Maria Barbosa. Introdução ao estudo do direito. Rio de Janeiro: América Jurídica, 2003.

RUSSELL, Stuart; NORVIG, Peter. Inteligência artificial. Tradução de Regina Célia Simille. 3. ed. Rio de Janeiro: Elsevier, 2013.

SANTOS, Jennifer Manfrin dos. A utilização da inteligência artificial no julgamento de casos simples para a promoção da celeridade processual nos juizados especiais cíveis. 2024. Dissertação (Mestrado em Direito) – Centro Universitário Internacional UNINTER, Curitiba, 2024. Disponível em: https://repositorio.uninter.com/handle/1/1721. Acesso em: 2 nov. 2026.

SIQUEIRA, Dirceu Pereira; MORAIS, Fausto Santos de; SANTOS, Marcel Ferreira dos. Inteligência artificial e jurisdição: dever analítico de fundamentação e os limites da substituição dos humanos por algoritmos no campo da tomada de decisão judicial. Seqüência: Estudos Jurídicos e Políticos, Florianópolis, v. 43, n. 91, p. 1-34, 2022. DOI: 10.5007/2177-7055.2022.e90662. Disponível em: https://periodicos.ufsc.br/index.php/sequencia/article/view/90662. Acesso em: 2 nov. 2026.

SUPREMO TRIBUNAL FEDERAL (Brasil). Inteligência artificial permitirá classificação dos processos do STF sob a ótica dos direitos humanos. Portal STF, Brasília, DF, 17 maio 2022. Disponível em: https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=487134&ori=1. Acesso em: 15 out. 2026.

SUSSKIND, Richard. Online courts and the future of justice. Oxford: Oxford University Press, 2019.

TRIBUNAL SUPERIOR DO TRABALHO. Robôs já estão presentes na maioria dos tribunais do país. [S. l.], 2022. Disponível em: https://portal.trt23.jus.br/trtnoticias/noticias/fique-sabendo-fique-por-dentro/28756. Acesso em: 2 nov. 2026.

WOTKOSKI, Daiane Medino. O papel da inteligência artificial na jurisdição contemporânea: utilização em decisões judiciais, eficiência e desafios no STF. 2024. Dissertação (Mestrado em Direito) – Centro Universitário Internacional UNINTER, Curitiba, 2024. Disponível em: https://repositorio.uninter.com/handle/1/1719. Acesso em: 15 out. 2026.

INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO: LIMITES E IMPACTOS NAS GARANTIAS PROCESSUAIS CIVIS

ARTIFICIAL INTELLIGENCE IN THE JUDICIARY: LIMITS AND IMPACTS ON CIVIL PROCEDURAL GUARANTEES

*Eliana da Silva Araújo**

RESUMO

Este artigo examina os limites ao uso de inteligência artificial pelo Poder Judiciário à luz das garantias processuais civis. A Resolução CNJ nº 615/2025, que substituiu a Resolução nº 332/2020, instituiu um regime de governança baseado em categorização de riscos, supervisão humana e transparência institucional, mas classificou como de baixo risco a produção de textos de apoio à elaboração de atos judiciais e tornou facultativa a menção ao uso de IA no corpo das decisões. Sustenta-se que a principal ameaça às garantias não está na decisão integralmente automatizada, vedada pela resolução, e sim na delegação de fato que se instala sob a forma de mero apoio e que não deixa vestígio no ato decisório. Analisam-se os efeitos dessa delegação sobre o contraditório substancial, o dever de fundamentação, a imparcialidade, a paridade de armas e a aplicação de precedentes. Conclui-se que a governança administrativa é necessária, porém insuficiente, e que a tutela das garantias depende de instrumentos internos ao processo: ciência das partes quanto ao emprego relevante de IA, contraditório prévio sobre fundamentos e precedentes de origem algorítmica e ônus argumentativo reforçado nas decisões apoiadas por esses sistemas.

Palavras-chave: Inteligência Artificial; Poder Judiciário; Garantias Processuais; Contraditório; Fundamentação das Decisões.

---

* Advogada com 34 anos de atuação. Docente do Ensino Superior há 19 anos. Mestre em Positivação e Concretização dos Direitos Humanos Fundamentais, pelo Centro Universitário FIEO - Fundação Instituto de Ensino para Osasco. Especialista em Direito Civil e Processo Civil, pelo Centro Universitário FIEO - Fundação Instituto de Ensino para Osasco.

ABSTRACT

This article examines the limits on the use of artificial intelligence by the Brazilian Judiciary in light of civil procedural guarantees. CNJ Resolution No. 615/2025, which replaced Resolution No. 332/2020, established a governance framework based on risk categorization, human oversight and institutional transparency, but classified the production of drafting aids for judicial acts as low risk and left disclosure of AI use in the text of decisions to the judge's discretion. The article argues that the main threat to procedural guarantees lies not in fully automated decision-making, which the resolution prohibits, but in de facto delegation that takes hold under the guise of mere support and leaves no trace in the decision itself. It analyzes the effects of such delegation on the adversarial principle, the duty to state reasons, impartiality, equality of arms and the application of precedents. It concludes that administrative governance is necessary but insufficient, and that the protection of procedural guarantees depends on instruments internal to the proceedings: disclosure to the parties of relevant AI use, a prior hearing on grounds and precedents of algorithmic origin, and a heightened burden of justification for decisions supported by such systems.

Keywords: Artificial Intelligence; Judiciary; Procedural Guarantees; Adversarial Principle; Duty to State Reasons.

1 INTRODUÇÃO

A presença de sistemas de inteligência artificial nos tribunais brasileiros deixou de ser experimental. Ferramentas de classificação e agrupamento de processos, de identificação de temas repetitivos, de pesquisa de precedentes e, mais recentemente, de redação assistida por modelos de linguagem passaram a integrar a rotina de gabinetes e secretarias. Referida expansão se confirma pelos dados da Pesquisa Inteligência Artificial no Poder Judiciário 2024, realizada pelo Conselho Nacional de Justiça (CNJ), que identificou 178 projetos de inteligência artificial em 58 tribunais, dos quais 60 já se encontravam em produção. O levantamento revelou, ainda, que 45,8% dos tribunais e conselhos participantes utilizavam ferramentas de inteligência artificial generativa, especialmente para geração, aprimoramento e sumarização de textos e documentos (CNJ, 2025).

A justificativa que acompanha essa expansão é conhecida e não é desprezível: um acervo de dimensões que a estrutura humana do Judiciário não absorve no tempo exigido pela garantia da razoável duração do processo (Art. 5º, inciso LXXVIII, da Constituição e Art. 4º do CPC).

O Conselho Nacional de Justiça respondeu a esse movimento com a Resolução nº 615, de 11 de março de 2025, que revogou a Resolução nº 332/2020 (Art. 46) e passou a disciplinar o desenvolvimento, a governança, a auditoria e o uso de soluções de IA no Poder Judiciário. O novo ato organiza-

se em torno de uma categorização de riscos de inspiração europeia, da exigência de supervisão humana e de deveres de transparência, e dedica capítulo próprio aos modelos de linguagem e aos sistemas de IA generativa. O prazo de doze meses concedido aos tribunais para adequar projetos e modelos (Art. 45, parágrafo único) já se esgotou, o que permite tomar a resolução não como promessa, mas como regime em vigor. E, como o Projeto de Lei n. 2338/2023, de autoria do senador Rodrigo Pacheco, que pretende instituir o marco legal da inteligência artificial no Brasil, o texto foi aprovado pelo Plenário do Senado Federal em dezembro de 2024, remetido à Câmara em março de 2025, ainda tramita na Comissão Especial sobre inteligência Artificial, Câmara dos Deputados. Portanto, a resolução n. 615/2025 é hoje o principal parâmetro normativo do uso jurisdicional dessas tecnologias.

O desenho adotado é, essencialmente, de governança institucional. Dirige-se aos tribunais como desenvolvedores e contratantes de sistemas e aos magistrados como usuários internos, impondo-lhes avaliação de risco, cadastro, auditoria e capacitação. Diz menos, contudo, sobre o que se passa no processo quando a ferramenta interfere no modo como o juiz compreende a causa. Dois dados do texto normativo evidenciam a lacuna: a produção de textos de apoio para a confecção de atos judiciais foi classificada como de baixo risco (Anexo, item BR4), e a menção ao emprego de IA no corpo da decisão ficou a critério do signatário (Art. 19, § 6º, e Art. 33, § 3º).

O problema que orienta este trabalho é saber em que medida o uso de IA pelos órgãos jurisdicionais, nos termos admitidos pela Resolução CNJ nº 615/2025, é compatível com as garantias que a Constituição e o Código de Processo Civil asseguram às partes, em especial o contraditório substancial, isto é, o direito de participar ativamente e de influenciar na decisão judicial, portanto, não é apenas tomar ciência dos atos processuais. A hipótese é a de que o risco mais sério não está na decisão inteiramente automatizada, que a resolução proíbe, mas na conversão gradual e pouco visível da ferramenta de apoio em fator determinante do julgamento. Essa delegação de fato ocorre antes do ato decisório e não deixa rastro nele: a decisão continua assinada por um magistrado e formalmente fundamentada, ainda que o percurso do convencimento tenha sido pautado por um resultado que as partes não conheceram nem puderam contestar. Se a hipótese estiver correta, a supervisão humana exigida pela regulação administrativa é condição necessária, mas não suficiente, e a tutela das garantias depende de instrumentos internos ao processo.

A pesquisa adota método dedutivo, com revisão bibliográfica e análise da Constituição, do CPC, da Resolução CNJ nº 615/2025 e, como contraponto, do Regulamento (UE) 2024/1689 e da Carta Ética Europeia sobre o uso da inteligência artificial nos sistemas judiciais. A seção 2 propõe uma tipologia funcional dos usos da IA no Judiciário e examina o regime da resolução. A seção 3 confronta essa tipologia com as garantias processuais afetadas, e a seção 4 formula critérios de compatibilização de natureza processual.

2 A INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO: FUNÇÕES E MARCO REGULATÓRIO

2.1 Uma tipologia funcional

A expressão "inteligência artificial no Judiciário" abrange atividades de natureza muito distinta, e a análise de seus efeitos sobre as garantias processuais fica comprometida quando a ferramenta que classifica petições iniciais e o modelo que redige uma minuta de sentença são tratados como manifestações do mesmo fenômeno. Para os fins deste artigo, propõe-se distinguir quatro níveis, ordenados segundo o grau de proximidade com o juízo decisório.

O primeiro nível compreende a gestão e a triagem: classificação processual, identificação de temas repetitivos, agrupamento de feitos semelhantes, extração de dados de documentos e execução de atos ordinatórios. O segundo abrange o apoio à pesquisa, notadamente a localização de precedentes e julgados por critérios de similaridade. O terceiro corresponde ao apoio cognitivo em sentido estrito, como o resumo de autos, a identificação de pontos controvertidos e a sugestão de enquadramento jurídico ou de precedente aplicável. O quarto é a produção de texto decisório, em que o sistema, em geral um modelo de linguagem, gera a minuta de despacho, decisão ou voto que o magistrado revisará.

A tipologia não pretende ser exaustiva nem tecnicamente rigorosa, mas demonstrar que uma mesma ferramenta pode transitar entre níveis diferentes conforme o uso que dela se faça. Seu propósito é evidenciar que, à medida que se avança do primeiro ao quarto nível, a ferramenta deixa de operar sobre o fluxo do processo e passa a operar sobre a formação do convencimento. É nesse deslocamento que as garantias processuais ficam expostas, razão pela qual a intensidade da proteção deve variar conforme o nível de sua aplicação.

2.2 O regime da Resolução CNJ nº 615/2025

A Resolução nº 615/2025 enuncia, entre seus princípios, o devido processo legal, a ampla defesa, o contraditório e a identidade física do juiz (art. 3º, inciso V), a explicabilidade e a contestabilidade das soluções (art. 3º, inciso II) e a supervisão humana efetiva, ajustável ao nível de automação e ao impacto da ferramenta (art. 3º, inciso VII). A contestabilidade prevista no art. 3º, inciso II, pode ser compreendida como a possibilidade de questionamento e revisão dos resultados gerados pela IA. E mais, ainda no art.3º, inciso VIII, atribui aos tribunais o dever de oferecer capacitação contínua sobre riscos da automação, vieses algorítmicos e análise crítica dos resultados.

A estrutura de riscos tem três patamares. O art. 10 veda, por risco excessivo, soluções que não admitam revisão humana ou que gerem dependência absoluta do usuário em relação ao resultado proposto, bem como as que valorem traços de personalidade para prever condutas, classifiquem pessoas para aferir a plausibilidade de seus direitos ou testemunhos ou reconheçam emoções a partir de padrões biométricos. As demais soluções são classificadas como de alto ou de baixo risco conforme o Anexo de Classificação de Riscos (art. 11). Entre as hipóteses de alto risco estão a valoração dos meios de prova quando possa influenciar diretamente a decisão (AR2) e a formulação de juízos conclusivos sobre a aplicação de norma jurídica ou de precedentes a um conjunto determinado de fatos concretos (AR4). Entre as de baixo risco figuram a sumarização de documentos e o agrupamento de processos (BR1), a detecção de padrões decisórios e de precedentes qualificados pertinentes (BR2) e a produção de textos de apoio para facilitar a confecção de atos judiciais, desde que a supervisão e a versão final caibam ao magistrado e se baseiem em suas instruções (BR4).

A fronteira entre AR4 e BR4 é o ponto mais delicado do regime. Uma minuta de sentença produzida por modelo de linguagem contém, por definição, um juízo sobre a aplicação do direito aos fatos da causa. O que a desloca para a categoria de baixo risco não é a natureza do conteúdo produzido, mas a circunstância de que o magistrado instrui a ferramenta e assina a versão final. A classificação repousa, portanto, inteiramente na eficácia da supervisão humana. Se essa supervisão for efetiva, a qualificação se sustenta; se for apenas formal, uma atividade materialmente de alto risco circulará sob o regime menos exigente, sem as auditorias regulares e a avaliação de impacto algorítmico reservadas às soluções de alto risco (arts. 11, § 1º, e 14).

O capítulo dedicado à IA generativa reforça essa centralidade. O art. 19 admite que magistrados e servidores utilizem modelos de linguagem como ferramentas de auxílio à gestão ou de apoio à decisão, preferencialmente por acesso disponibilizado e monitorado pelo tribunal, e autoriza, na falta de solução corporativa, a contratação direta de serviços privados (§§ 1º e 2º). O uso deve ser auxiliar e complementar, vedada a utilização como instrumento autônomo de decisão, e o magistrado permanece integralmente responsável pelas decisões e pelas informações nelas contidas (§ 3º, II); ferramentas externas ao Judiciário não podem ser empregadas em finalidades de alto risco (§ 3º, V). No plano da transparência, o § 6º determina o registro automático, no sistema interno do tribunal, do emprego de IA generativa na redação de atos judiciais, para fins estatísticos, de monitoramento e de auditoria, mas deixa a menção no corpo da decisão a critério do magistrado. O art. 33, § 3º, reitera que essa comunicação é faculdade do signatário.

O capítulo sobre o controle do usuário completa o quadro. O sistema deve permitir a revisão detalhada do conteúdo gerado e dar acesso às premissas e ao método empregados, sem vinculação à solução proposta (art. 32, II), e não pode, em nenhum momento, restringir ou substituir a autoridade final dos usuários internos (art. 32, parágrafo único). Os usuários externos, categoria que inclui advogados e jurisdicionados (art. 4º, VII), devem ser informados sobre a utilização de IA nos serviços que lhes são prestados, com destaque para o caráter consultivo e não vinculante da proposta gerada (art. 33, caput e § 1º). Essa informação, contudo, circula por avisos, materiais informativos e guias (art. 33, § 2º), e não necessariamente no processo em que a ferramenta foi usada.

O resultado é um regime que conhece bem o risco da automação explícita e o enfrenta com vedações e controles consistentes, mas que concebe a transparência sobretudo como prestação de contas institucional: catálogos públicos, relatórios, registros internos e acesso da OAB, do Ministério Público e da Defensoria Pública às avaliações de impacto e aos relatórios de auditoria (arts. 5º, § 3º, e 12, VIII). As partes de um processo concreto não figuram como destinatárias de informação sobre o que a ferramenta fez naquele processo.

2.3 Regulação administrativa de matéria processual

A circunstância de o uso jurisdicional da IA estar disciplinado por resolução do CNJ não é neutra. O PL 2.338/2023, aprovado pelo Senado Federal em dezembro de 2024, encontra-se em comissão especial na Câmara dos Deputados, e não há, por ora, lei que trate da matéria sob a perspectiva do

processo. O CNJ atua no exercício do controle da atuação administrativa do Poder Judiciário (art. 103-B, § 4º, da Constituição), ao passo que a legislação sobre direito processual é de competência privativa da União (art. 22, I). Daí resulta uma limitação estrutural: a resolução pode disciplinar como os tribunais desenvolvem, contratam, auditam e disponibilizam ferramentas, mas não pode criar, em favor das partes, direitos processuais exigíveis no processo, nem cominar nulidades. Acrescente-se que o Supremo Tribunal Federal não se submete ao controle do CNJ, (ADI 3.367, julgada 13 de abril de 2005), de modo que o regime não alcança o uso de IA na Corte.

Essa limitação não é um defeito a ser corrigido pela própria resolução; é antes uma indicação de onde procurar as respostas. Se as garantias atingidas são processuais, a proteção também deve vir das normas processuais existentes, interpretadas à luz do novo contexto tecnológico. O argumento das seções seguintes é que o CPC de 2015 já oferece os instrumentos necessários, desde que se reconheça que a interferência da IA na formação do convencimento é um fato processualmente relevante.

3 GARANTIAS PROCESSUAIS CIVIS SOB PRESSÃO ALGORÍTMICA

3.1 Contraditório substancial e vedação à decisão surpresa

O CPC de 2015 consagrou o contraditório como direito de influência. Não basta que a parte seja ouvida; é preciso que tenha oportunidade efetiva de participar da construção da decisão, o que se traduz na vedação de decisão contra a parte sem sua prévia oitiva (art. 9º) e na proibição de decidir com base em fundamento a respeito do qual não se tenha dado às partes oportunidade de manifestação, ainda que se trate de matéria apreciável de ofício (art. 10). O art. 7º, por sua vez, impõe ao juiz o dever de zelar pelo efetivo contraditório.

Os níveis 2 e 3 da tipologia tensionam diretamente essa garantia. Quando uma ferramenta de pesquisa por similaridade localiza um precedente que nenhuma das partes invocou, ou quando um sistema de apoio cognitivo sugere enquadramento jurídico diverso daquele debatido nos autos, o magistrado passa a dispor de um elemento possivelmente decisivo que as partes desconhecem. Não há, nisso, novidade em relação à pesquisa jurisprudencial feita pela assessoria, e o art. 10 sempre exigiu que fundamentos novos fossem submetidos às partes, qualquer que fosse sua origem. A diferença está na escala e na invisibilidade. A ferramenta

produz sugestões em volume e com aparência de neutralidade técnica, e o registro interno de seu uso, quando existe, não chega ao conhecimento dos litigantes. A decisão surpresa de origem algorítmica é, por isso, mais provável e mais difícil de detectar.

No nível 4 o problema se agrava. A minuta gerada por modelo de linguagem pode incorporar argumentos, qualificações jurídicas e até referências que não constavam dos autos, e o magistrado que a revisa tende a avaliar sua coerência interna, em vez de confrontá-la com o que foi efetivamente debatido. O contraditório substancial pressupõe que aquilo que influencia o juiz esteja ao alcance das partes. Quando a influência decisiva vem de um texto que elas não viram, a garantia permanece íntegra no plano formal e esvazia-se no plano material.

3.2 Dever de fundamentação

A exigência constitucional de fundamentação das decisões (art. 93, IX) ganhou densidade no art. 489, § 1º, do CPC, que considera não fundamentada a decisão que se limita a reproduzir ato normativo, emprega conceitos indeterminados sem explicar sua incidência, invoca motivos que se prestariam a justificar qualquer outra decisão, deixa de enfrentar os argumentos capazes de infirmar a conclusão ou aplica precedente sem identificar seus fundamentos determinantes e demonstrar o ajuste ao caso. A omissão nesses pontos autoriza a oposição de embargos de declaração (art. 1.022, parágrafo único, II).

A IA generativa afeta essa garantia de dois modos. O primeiro é visível e diz respeito à qualidade do texto. Modelos de linguagem produzem com facilidade fundamentações plausíveis, bem articuladas e genéricas, isto é, exatamente o tipo de motivação que o § 1º, inciso II, do art. 489, reputa inexistente. A fluência do texto dissimula a ausência de enfrentamento específico dos argumentos das partes, e a revisão apressada tende a não perceber o que falta. Somam-se a isso os casos de referências inexistentes ou distorcidas produzidas por esses sistemas, risco que a própria resolução procura conter ao exigir que as soluções contratadas ofereçam documentação e referências bibliográficas atualizadas (art. 20, VIII).

Destacando que o ministro da 6ª turma do Superior Tribunal de Justiça, Rogerio Schietti Cruz, identifica *habeas corpus* com erros e alucinações criadas por IA, que geraram referências erradas a precedentes judiciais e trechos de julgados inexistentes. Após ouvir o advogado signatário da peça, determinou que o caso fosse comunicado a OAB, para as providências cabíveis.

Salientando que na decisão que negou a liminar, o ministro Schietti manifestou-se no sentido de que "O uso da inteligência artificial na prática jurídica não é, em si, censurável. Trata-se de recurso que, quando bem empregado, pode qualificar o trabalho advocatício e racionalizar o esforço judicial. O problema está na ausência de verificação humana do conteúdo gerado. A tecnologia serve ao profissional, mas não o substitui nem o desobriga de conferir o que assina", afirmou Schietti.

O segundo modo é menos visível e mais grave. A fundamentação cumpre função de controle porque expõe as razões pelas quais o juiz decidiu. Quando a minuta é gerada por uma ferramenta e revisada pelo magistrado, as razões expostas podem ser apenas as que o sistema articulou, enquanto o processo real de convencimento, a escolha entre aceitar ou alterar a proposta, permanece fora do texto. A decisão pode estar formalmente fundamentada e, ainda assim, ocultar o caminho efetivo pelo qual se chegou a ela. A opacidade técnica dos modelos, que a resolução procura mitigar por exigências de explicabilidade "sempre que tecnicamente possível" (arts. 13, VII, e 22, § 3º), agrava essa dissociação entre a decisão justificada e a decisão efetivamente tomada (PASQUALE, 2015).

3.3 Imparcialidade e viés de automação

A imparcialidade costuma ser pensada a partir dos vínculos do juiz com as partes ou com o objeto da causa. O uso de sistemas de IA introduz outra dimensão, de ordem cognitiva. A literatura sobre a interação entre pessoas e sistemas automatizados descreve o viés de automação como a tendência de tomar a recomendação do sistema como atalho heurístico, deixando de buscar ou de valorizar informações contrárias e de perceber erros que o sistema não sinalizou (SKITKA; MOSIER; BURDICK, 1999; PARASURAMAN; MANZEY, 2010). Os estudos indicam que o efeito não desaparece com a experiência do usuário e tende a se intensificar sob alta carga de trabalho.

A transposição para o contexto judicial é imediata. O magistrado que revisa dezenas de minutas por dia, sob pressão de metas de produtividade, está precisamente na condição em que o viés de automação mais se manifesta. A supervisão humana exigida pela resolução existe juridicamente, mas pode degradar-se em ratificação. O próprio art. 10, inciso I, ao vedar soluções que gerem dependência absoluta do usuário, reconhece o fenômeno. A dificuldade é que a dependência relevante não decorre apenas do desenho da ferramenta, mas das condições concretas em que ela é utilizada, e uma ferramenta que permite revisão plena pode ser usada sem revisão efetiva.

A 4ª Turma do STJ ao examinar alegação de uso de IA em decisão judicial, AgInt no AREsp 2.848.833/PR, julgado em 15 de junho de 2026 e publicado em 22 de junho, por unanimidade negou provimento ao recurso, afastando nulidade fundada em mera suspeita, suscitando questões sobre fundamentação, supervisão humana e regime das invalidades.

Destacando que no exame desse julgado verifica-se que o STJ não descarta o uso de IA, porque apenas analisou suspeitas alegadas pelos recorrentes de que a decisão carecia de fundamentação particular ou própria e que ainda "teria sido proferida por inteligência artificial, sem revisão humana efetiva", dessa maneira, violou o devido processo legal, ao contraditório, à ampla defesa e à Resolução do CNJ n. 615/25.

A ementa desse acórdão, em conclusão, sintetizou não reconhecer nulidade apontada "com base em alegações genéricas de uso de ferramenta de inteligência artificial, sem demostração concreta de prolação automática, ausência de supervisão humana ou prejuízo processual efetivo".

Há, ainda, o problema dos vieses dos dados. Sistemas treinados com decisões passadas tendem a reproduzir os padrões nelas presentes, inclusive os que o ordenamento pretende superar. A resolução enfrenta o ponto ao exigir dados representativos e curadoria (arts. 7º e 26) e ao prever a descontinuação de soluções com viés discriminatório insuperável (art. 8º, § 3º). No processo civil, porém, o risco mais frequente não é a discriminação de grupos, mas a cristalização da orientação jurisprudencial majoritária, que a ferramenta passa a sugerir como resposta padrão. O juiz tem mais dificuldade para perceber as particularidades do caso, e a jurisprudência, para evoluir.

3.4 Paridade de armas e publicidade

O art. 7º do CPC assegura às partes paridade de tratamento no exercício de faculdades processuais. A difusão de ferramentas de IA cria uma assimetria que as regras processuais não previram. Litigantes habituais, como grandes empresas, instituições financeiras e o próprio poder público, dispõem de sistemas capazes de mapear padrões decisórios de juízos e órgãos colegiados e de calibrar suas estratégias em função deles, enquanto o litigante eventual não tem acesso a nada semelhante. Quando o próprio tribunal opera com ferramentas de detecção de padrões decisórios (Anexo, itens BR2 e BR3), forma-se um circuito em que os repetidores de demandas conhecem melhor o funcionamento provável da decisão do que a parte que litiga uma única vez.

A publicidade dos atos processuais (arts. 5º, LX, e 93, IX, da Constituição; art. 11 do CPC) também é tensionada. A resolução privilegia a publicidade institucional, por meio do catálogo de aplicações (art. 25) e da indicação, na interface dos sistemas de processo eletrônico, dos modelos em uso (art. 21). Essas medidas informam que o tribunal emprega determinadas ferramentas, mas não se e como elas foram empregadas em um processo específico. A parte sabe que a IA existe no tribunal; não sabe se ela interferiu na sua causa.

3.5 Precedentes e distinção

O sistema de precedentes instituído pelos arts. 926 e 927 do CPC depende de uma operação essencialmente comparativa e fática: identificar os fundamentos determinantes do precedente e demonstrar que o caso sob julgamento a eles se ajusta ou, ao contrário, deles se distingue (art. 489, § 1º, V e VI). Ferramentas de recuperação e agrupamento operam, em regra, por similaridade textual ou semântica entre documentos, e essa similaridade não coincide com a identidade de fundamentos determinantes. Dois casos podem ser descritos com vocabulário quase idêntico e diferir justamente no fato que torna inaplicável a tese firmada.

Um exemplo hipotético ilustra o ponto. Em demandas sobre cobrança de tarifas bancárias, a ferramenta agrupa centenas de processos sob o mesmo tema repetitivo porque as petições iniciais compartilham pedidos e vocabulário. Em um deles, porém, o autor alega que a tarifa foi cobrada sem contratação expressa, circunstância que, por hipótese, não foi considerada no julgamento que fixou a tese e pode afastar sua aplicação. Se a minuta padronizada do grupo for adotada, a alegação específica não será enfrentada, e a decisão incorrerá no vício do art. 489, § 1º, IV e VI, sem que nada em seu texto revele que a distinção sequer foi considerada. O exemplo é deliberadamente simples, e os casos reais tendem a ser menos nítidos, o que só agrava o problema.

A resolução classifica como de baixo risco a detecção de precedentes qualificados pertinentes, desde que observado o caráter complementar da técnica e não haja substituição da avaliação humana (Anexo, item BR2), e como de alto risco a formulação de juízos conclusivos sobre a aplicação de precedentes a fatos concretos (AR4). Na prática, a passagem de uma hipótese à outra é tênue. O agrupamento automático de processos sob um tema repetitivo, seguido de minutas padronizadas que aplicam a tese, produz exatamente o efeito que o art. 489, § 1º, VI, pretende evitar: a aplicação do precedente por semelhança superficial, sem enfrentamento

do argumento de distinção. O risco aumenta porque a parte que pretende demonstrar a distinção litiga contra um fluxo de trabalho desenhado para tratar seu processo como mais um do grupo.

4 LIMITES PROCESSUAIS AO USO DE INTELIGÊNCIA ARTIFICIAL

4.1 A formação do juízo como núcleo indelegável

A afirmação de que a decisão judicial deve permanecer humana é hoje consensual e está expressa na resolução (arts. 19, § 3º, II, 32, parágrafo único, e 34). O que se sustenta aqui é que essa reserva não se satisfaz com a autoria formal do ato. Parte da literatura recente tem tratado a questão sob a ideia de uma reserva de humanidade. Para que tenha conteúdo, a reserva precisa alcançar a formação do juízo: a apreciação das provas (art. 371 do CPC), a interpretação das normas e a escolha entre as soluções juridicamente possíveis. Não por acaso, são essas as atividades que o próprio Anexo trata como de alto risco quando desempenhadas ou apoiadas por IA (AR2 e AR4).

Daí decorre um critério interpretativo para a fronteira entre AR4 e BR4. A produção de texto de apoio só pode ser tratada como de baixo risco quando se limita a dar forma a um juízo previamente formado pelo magistrado, com base em suas instruções, como exige a própria redação do item BR4. Quando o sistema propõe o juízo e o magistrado se limita a aceitá-lo, ainda que com ajustes, a atividade é materialmente de alto risco, e o uso de ferramentas externas ao Judiciário para esse fim é vedado (art. 19, § 3º, V). A distinção não é meramente classificatória, pois define se a supervisão humana foi exercida sobre a decisão ou apenas sobre o texto.

4.2 Do registro interno à ciência das partes

A resolução impõe o registro do uso de IA generativa no sistema interno do tribunal (art. 19, § 6º) e a identificação, nos registros de uso do sistema, dos produtos gerados de forma automatizada (art. 21, § 2º), mas deixa ao magistrado decidir se a decisão mencionará o emprego da ferramenta (art. 33, § 3º). Do ponto de vista da governança, a opção é compreensível: o registro serve à auditoria e à estatística, e a menção obrigatória poderia, na lógica do art. 1º, § 3º, afetar a percepção de credibilidade das decisões. Do ponto de vista processual, porém, a solução priva as partes da única informação que lhes permitiria exercer o contraditório sobre o que a ferramenta produziu.

Essa faculdade não pode ser lida isoladamente. Por se tratar de norma administrativa, ela não afasta os deveres que o CPC impõe ao juiz. Quando o uso da IA se limitar aos níveis 1 e 2 sem introduzir fundamento novo, ou a tarefas de mera formatação (Anexo, item BR5), a menção é de fato dispensável. Quando, porém, a ferramenta tiver contribuído para a seleção de fundamento ou precedente não debatido, ou para a construção do juízo sobre os fatos e o direito, a ciência das partes deixa de ser faculdade e passa a ser exigência do contraditório (arts. 9º e 10) e do dever de fundamentação, que compreende a exposição do caminho efetivamente percorrido. Propõe-se, assim, uma interpretação conforme: a faculdade do art. 33, § 3º, opera apenas no espaço que as garantias processuais deixam livre.

A objeção previsível é a de que essa exigência sobrecarregaria a atividade jurisdicional e anularia os ganhos de eficiência que justificam o uso da tecnologia. Ela procede apenas em parte. A ciência das partes não exige relatório técnico nem exposição do funcionamento do modelo; basta a indicação de que determinada ferramenta foi utilizada para determinada finalidade naquele ato. O ônus é pequeno quando comparado ao custo de decisões posteriormente anuladas ou reformadas por violação ao contraditório. E a própria resolução já impõe o registro automático do uso (art. 19, § 6º): o que se propõe é que a informação já produzida para fins de auditoria seja acessível às partes do processo em que foi gerada, e não apenas ao tribunal.

De lege ferenda, seria desejável que essa exigência fosse positivada. O marco legal em discussão no Congresso, ou uma alteração pontual do CPC, poderia prever o dever de indicar nos autos o emprego de sistemas de IA nas atividades correspondentes aos níveis 3 e 4, com identificação da ferramenta e da finalidade, e o direito da parte de requerer acesso ao resultado gerado quando relevante para a decisão.

4.3 Contraditório sobre o resultado algorítmico

A segunda medida decorre diretamente do art. 10 do CPC. Se o sistema sugerir fundamento, qualificação jurídica ou precedente que não foi objeto de debate e que o magistrado pretenda adotar, as partes devem ser intimadas para manifestação antes da decisão. Não se trata de instaurar contraditório sobre o funcionamento interno do modelo, o que seria impraticável, mas sobre o seu resultado, isto é, sobre o elemento que ingressará na decisão.

A inobservância dessa exigência conduz ao regime das nulidades. A decisão que adota fundamento de origem algorítmica sem prévia oitiva das partes viola o art. 10 do CPC nos mesmos termos de qualquer decisão

surpresa, e incide o filtro da instrumentalidade e do prejuízo (arts. 277 e 282, §§ 1º e 2º), de modo que a nulidade será reconhecida quando o fundamento não debatido tiver sido relevante para o resultado. O que muda é o problema da prova. Como a parte não tem acesso ao registro interno do uso de IA, a demonstração do vício dependerá muitas vezes de indícios no próprio texto decisório, como referências inexistentes, fundamentos alheios aos autos ou o enfrentamento de argumentos que ninguém formulou. Também por isso a ciência prévia defendida no item anterior é relevante.

4.4 Ônus argumentativo reforçado

A terceira medida incide sobre a fundamentação. A decisão produzida com apoio de IA nos níveis 3 e 4 deve satisfazer o art. 489, § 1º, com rigor acrescido em dois pontos. Deve enfrentar expressamente os argumentos das partes que contrariem a solução adotada, pois é aí que a revisão de um texto gerado tende a falhar: o modelo organiza a fundamentação em torno da conclusão e omite o que a infirma. E, na aplicação de precedentes, deve demonstrar a correspondência entre os fatos do caso e os fundamentos determinantes do paradigma, e não apenas a semelhança temática que orientou a recuperação ou o agrupamento automático.

Esse ônus reforçado não depende de inovação normativa. Resulta da leitura do art. 489, § 1º, à luz das condições concretas em que a decisão é produzida, e pode ser exigido pelas partes por meio de embargos de declaração por omissão (art. 1.022, parágrafo único, II) e do controle recursal. Para os tribunais, a consequência prática é que a capacitação prevista nos arts. 3º, VIII, e 19, § 5º, da resolução deve ir além do manejo das ferramentas e abranger os deveres processuais que seu uso torna mais exigentes.

4.5 Um contraponto europeu

A experiência europeia revela convergência quanto ao diagnóstico, mas não resolve o problema processual aqui identificado. A Carta Ética Europeia sobre o uso da inteligência artificial nos sistemas judiciais, adotada pela CEPEJ em 2018, já enunciava entre seus princípios o respeito aos direitos fundamentais e o controle pelo usuário, segundo o qual o juiz deve poder rever as decisões e os dados utilizados e as partes devem ser informadas sobre o caráter vinculante ou não das soluções propostas (CEPEJ, 2018). O Regulamento (UE) 2024/1689 classifica como de alto risco os sistemas

destinados a auxiliar autoridades judiciárias na pesquisa e interpretação de fatos e do direito e na aplicação do direito a um conjunto concreto de fatos, e afirma que a decisão final deve permanecer uma atividade humana.

A resolução brasileira dialoga visivelmente com esse modelo, inclusive na técnica de classificar finalidades por risco. A diferença relevante está na qualificação da redação assistida: o que o regulamento europeu tende a tratar como alto risco quando o sistema auxilia a aplicação do direito aos fatos, a resolução brasileira admite como baixo risco quando a versão final cabe ao magistrado (BR4). Em ambos os casos, porém, o foco recai sobre obrigações de quem desenvolve e de quem implanta os sistemas. Nenhum dos dois instrumentos converte essas obrigações em direitos processuais das partes no caso concreto, tarefa que permanece a cargo do direito processual de cada ordenamento.

5 CONSIDERAÇÕES FINAIS

O exame da Resolução CNJ nº 615/2025 mostra que o Judiciário brasileiro dispõe de um regime de governança da inteligência artificial tecnicamente cuidadoso, que veda as formas mais evidentes de automação decisória e impõe deveres consistentes de avaliação de risco, auditoria e transparência institucional. O regime, contudo, foi construído para controlar sistemas, e não processos. Seu principal pressuposto, a supervisão humana, é tratado como garantia suficiente, e é sobre ele que repousa a classificação como de baixo risco da produção de textos de apoio à decisão.

A hipótese da pesquisa se confirma nesse ponto. A ameaça mais relevante às garantias processuais civis está na delegação de fato que se instala sob a forma de apoio: o fundamento que a ferramenta sugere e as partes não discutem, a minuta que o magistrado ratifica sem reconstruir o juízo, o precedente aplicado por semelhança de superfície. Nenhum desses fenômenos deixa vestígio no ato decisório, e nenhum é alcançado pelos controles institucionais, que operam fora do processo.

A resposta, por isso, precisa ser processual. Propôs-se que a faculdade de mencionar o uso de IA seja interpretada à luz dos arts. 9º e 10 do CPC, convertendo-se em exigência quando a ferramenta tiver contribuído para fundamento ou juízo relevante; que o resultado algorítmico decisivo e não debatido seja submetido a contraditório prévio, sob pena de nulidade condicionada ao prejuízo; e que as decisões apoiadas por IA cumpram o art. 489, § 1º, com ônus argumentativo reforçado. Essas medidas não

dependem de reforma legislativa, embora a positivação de um dever de indicação nos autos e de um direito de acesso ao resultado relevante seja desejável no marco legal em tramitação. O que está em jogo é que a decisão judicial continue sendo, também em sua formação, um ato construído com as partes.

Permanecem em aberto questões que este artigo apenas tangenciou e que merecem investigação própria. A primeira é empírica: em que medida os registros internos previstos na resolução permitem reconstruir, em cada processo, a contribuição efetiva da ferramenta para a decisão, e se esses registros têm sido produzidos de modo uniforme pelos tribunais. A segunda diz respeito ao acesso das partes aos registros de uso, à luz da Lei de Acesso à Informação e da Lei Geral de Proteção de Dados. A terceira concerne ao alcance do regime em relação ao Supremo Tribunal Federal, que permanece fora do controle do CNJ e cujas decisões têm, no sistema de precedentes, peso singular.

REFERÊNCIAS BIBLIOGRÁFICAS

BORBA, Alderico Kleber de. **O Direito de Influência no julgamento no Código de Processo Civil: contraditório tridimensional.** 1ª edição. Belo Horizonte: Editora Conhecimento, 2018

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

BRASIL. Lei nº 13.105, de 16 de março de 2015. **Código de Processo Civil**. Brasília, DF: Presidência da República, 2015.

BRASIL. Congresso Nacional. Senado Federal. Projeto de Lei nº 2338, de 2023. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília: Senado Federal, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 2 nov. 2026.

BRASIL. CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 615, de 11 de março de 2025**. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 24 out. 2026.

BRASIL. Conselho Nacional de Justiça. **Pesquisa Inteligência Artificial (IA) no Poder Judiciário**. Brasília: CNJ, 2025. Disponível em: https://www.cnj.jus.br/ia-generativa-e-utilizada-em-mais-de-45-dos-tribunais-brasileiros. Acesso em: 2 nov. 2026.

BRASIL. Superior Tribunal de Justiça. O dever de fundamentação e o uso de inteligência artificial na redação de decisões judiciais. Revista de Estudos Jurídicos do Superior Tribunal de Justiça (REJURI), Brasília, v. 4, 2026. Disponível em: stj.jus.br. Acesso em: 2 nov. 2026.

BRASIL. Superior Tribunal de Justiça. Habeas Corpus n. 1094270-MG 2026. Relator Ministro Rogério Schietti Cruz, 6ª turma, julgado em 18.05.2026, publicado no DJEN em 20.05.2026. Acesso em: 04.11.2026

BRASIL. Superior Tribunal de Justiça (4.ª Turma). Agravo Interno no Agravo em Recurso Especial n. 2.848.833/PR. Relatora: Min. Maria Isabel Gallotti, sessão virtual de 09 a 15 de junho de 2026. Diário da Justiça Eletrônico Nacional (DJEN), 22 jun. 2026.

CONSELHO DA EUROPA. Comissão Europeia para a Eficiência da Justiça (CEPEJ). Diretrizes sobre Inteligência Artificial na Administração da Justiça. Estrasburgo: CEPEJ, 2025. Disponível em: https://www.coe.int/en/web/cepej. Acesso em: 2 nov. 2026.

DIDIER JR., Fredie; CUNHA, Leonardo Carneiro da. **Tecnologia e Processo**. Salvador: Juspodivm, 2024.

GABRIEL, Anderson de Paiva; PORTO, Fabio Ribeiro; ARAÚJO, Valter Shuenquener de. Inteligência artificial e a Resolução CNJ 615/2025. **Revista da EMERJ**, *[S. l.]*, v. 27, p. 1–17, 2025. DOI: 10.70622/2236-8957.2025.649. Disponível em: https://ojs.emerj.com.br/index.php/revistadaemerj/article/view/649. Acesso em: 2 nov. 2026.

MARINONI, Luiz Guilherme. Precedentes obrigatórios. 7ª edição. São Paulo: Revista dos Tribunais, 2022.

MARINONI, Luiz Guilherme; ARENHART, Sérgio Cruz. Prova e convicção. 6ª edição. São Paulo: Revista dos Tribunais, 2022.

MARINONI, Luiz Guilherme; MITIDIERO, Daniel. **O Estatuto dos Precedentes e a Inteligência Artificial**. São Paulo: Thomson Reuters Revista dos Tribunais, 2025.

PASQUALE, Frank. **The black box society**: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015. https://doi.org/10.4159/harvard.9780674736061. Acesso: 02 de novembro de 2026.

SKITKA, Linda J.; MOSIER, Kathleen L.; BURDICK, Mark. Does automation bias decision-making?. International Journal of Human-Computer Studies, v. 51, n. 5, p. 991-1006, 1999. Disponível em: https://doi.org/10.1006/ijhc.1999.0252. Acesso em: 02 de nov de 2026.

TJDFT - Tribunal de Justiça do Distrito Federal e dos Territórios. Fundamentação das decisões judiciais. Tema criado em 30/3/2023, última modificação 09/06/2023.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024. Estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e que altera os Regulamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) n.o 2018/858, (UE) n.o 2018/1139 e (UE) n.o 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828. Jornal Oficial da União Europeia, Bruxelas, série L, 12 jul. 2024. Disponível em: https//:eur-lex.europa.eu/legal-content/PT/TXT/?uri=OJ:L_202401689. Acesso em: 2 nov. 2026.

WAMBIER, Teresa Arruda Alvim. **Nulidades do Processo e da Sentença**. 11ª edição. São Paulo: Revista dos Tribunais, 2022

Crimes Cibernéticos e o Golpe do Falso Advogado e Binance

Cybercrimes and the Fake Lawyer Scam and Binance

*Mário Sérgio de Oliveira**

RESUMO

O presente artigo analisa a evolução dos crimes cibernéticos e os desafios enfrentados pelo ordenamento jurídico brasileiro diante do avanço das tecnologias digitais. O estudo tem como objetivo examinar as principais modalidades de delitos praticados no ambiente virtual, com destaque para o golpe do falso advogado e os ataques envolvendo criptoativos, especialmente o caso da corretora Binance. Por meio de pesquisa bibliográfica e documental, são abordados o desenvolvimento histórico da criminalidade cibernética, suas consequências sociais e econômicas e as principais inovações legislativas destinadas à prevenção e repressão dessas práticas. Discute-se, ainda, a utilização da inteligência artificial na sofisticação das fraudes digitais, especialmente mediante técnicas de deepfake e falsificação de identificadores de chamadas. A análise evidencia que, embora o Brasil tenha ampliado seus mecanismos jurídicos de enfrentamento aos crimes cibernéticos, persistem dificuldades relacionadas à identificação dos responsáveis, à recuperação de valores e à atuação de organizações criminosas transnacionais. Conclui-se que a efetividade do combate à criminalidade digital exige não apenas o aperfeiçoamento legislativo, mas também investimentos em segurança tecnológica, medidas preventivas e fortalecimento da cooperação internacional.

Palavras-chave: Crimes cibernéticos; Fraude eletrônica; Falso advogado; Inteligência artificial; Criptoativos.

---

* Advogado criminal, ex-relator do Tribunal de Ética da OAB/ SP, ex-Coordenador da Comissão de prerrogativas da OAB/SP, Láurea de professor oferecida pela OAB/SP, professor desde 2002, atual vice-presidente da Comissão de Direito Penal da subseção de Cotia, palestrante da OAB/SP

ABSTRACT

This article analyzes the evolution of cybercrime and the challenges faced by the Brazilian legal system in response to advances in digital technologies. The study aims to examine the main types of offenses committed in virtual environments, focusing on the fake lawyer scam and attacks involving cryptoassets, particularly the case of the Binance cryptocurrency exchange. Based on bibliographic and documentary research, the article addresses the historical development of cybercrime, its social and economic consequences, and the main legislative innovations aimed at preventing and combating such practices. It also discusses the use of artificial intelligence to increase the sophistication of digital fraud, particularly through deepfake techniques and caller ID spoofing. The analysis demonstrates that, although Brazil has expanded its legal mechanisms to combat cybercrime, significant challenges remain regarding the identification of perpetrators, the recovery of stolen assets, and the activities of transnational criminal organizations. The study concludes that effectively combating digital crime requires not only legislative improvements but also investments in technological security, preventive measures, and stronger international cooperation.

Keywords: Cybercrime; Electronic fraud; Fake lawyer scam; Artificial intelligence; Cryptoassets.

1 INTRODUÇÃO

Com a evolução da internet, a qual teve seu início nos anos 60, criada nos Estados Unidos pelo Departamento de Defesa, ocorrendo a primeira conexão bem sucedida em 29 de outubro de 1969, surge em 1972 o e-mail, nos finais dos anos 1980, chega ao Brasil as primeiras conexões, no ano de 1994/1995 ocorre a liberação comercial da internet no país.

Com isso os criminosos passam a se utilizar dessa nova tecnologia, mantendo o anonimato, para manipulação de dados bancários, piratarias, pornografia infantil, migrando posteriormente para todo tipo de golpes em larga escala, incluindo a advocacia, com o famoso golpe do falso advogado, a cada dia mais aperfeiçoado, o qual será abordado mais adiante neste artigo. (Agência Câmara)

Hoje temos verdadeiras empresas praticando golpes digitais, citando por exemplo o ataque hacker que permitiu, até o presente momento, o maior roubo de criptomoedas já conhecido. A corretora Binance sofreu um roubo no valor aproximado de US$ 6 bilhões em criptomoedas.

Mas não somente as grandes instituições sofrem ataques pela internet, “Quem nunca recebeu uma mensagem suspeita? Quem não conhece alguém que já foi vítima de uma fraude cibernética?”

Desde 2018 o crescimento foi de **430%** de golpes pela internet, superando o crime de roubo como principal crime patrimonial do país.

A impunidade também é um incentivo aos criminosos, mais de 97% dos casos não chegam ao Judiciário, seja porque a vítima não acreditar numa punição, ou seja, por vergonha de confessar ter caído num golpe.

No ano de 2025 o Brasil registrou **2,26 milhões de estelionatos** — 258 por hora, mais de quatro por minuto; (Fonte**:** 20º Anuário Brasileiro de Segurança Pública, FBSP, 2026).

Diante dessa situação, em que o "O ladrão de hoje não precisa de arma. Precisa só que você acredite nele por alguns minutos", a Legislação Penal Brasileira, apesar de já ter normas para algumas dessas condutas criminosas, se viu obrigada a criar novas regras para punir os infratores de crimes praticados através do celular, computador utilizando a internet como meio.

Esses crimes podem ser praticados atacando o sistema em si — invasão de dispositivo, ransomware, chamados de ataques próprios, ou impróprios, crimes já conhecidos como venda de roupa que não são entregues, estelionato, ameaça, ofensas, perseguição entre outras.

Para se adaptar a esses novos tempos novas leis surgiram:

- 2012 - Lei Carolina Dieckmann: invasão de dispositivo vira crime.
- 2014 - Marco Civil da Internet: guarda de registros e regras para plataformas.
- 2021 - Lei 14.155: cria a fraude eletrônica; Lei 14.132: crime de stalking.
- 2024 - Lei 14.811: cyberbullying vira crime.
- 2025 - Lei 15.123: pena maior para violência psicológica contra a mulher com uso de IA (deepfakes); ECA Digital protege crianças online.
- **Lei 15.397/2026** (em vigor desde 04/05/2026):
  - Fraude eletrônica: **4 a 8 anos** de reclusão.
  - Novo crime: **ceder conta laranja** - 1 a 5 anos. ("Emprestar a conta para um conhecido receber dinheiro agora é crime."
  - Furto de celular ou computador: 4 a 10 anos.
  - Estelionato volta a ser de **ação penal incondicionada**: o MP processa mesmo sem representação da vítima.

Os golpes mais comuns são:

- **Falsa central do banco:** “compra suspeita” → Pix para “conta segura” ou motoboy buscando o cartão.
- **WhatsApp clonado ou “número novo”:** pedem o código de verificação ou usam sua foto para pedir dinheiro à família.
- **Phishing:** links falsos por SMS, e-mail ou anúncio - foram **22.818 notificações** em 2025 (CERT.br).
- **Lojas, leilões e investimentos falsos:** preço bom demais, lucro garantido, criptomoeda.
- **Falso emprego e “tarefas pagas”:** ganha pouco no início e depois é levado a “investir”.
- **Sextorsão e deepfakes:** ameaça de divulgar imagens íntimas, reais ou montadas por IA.
- **Crimes contra a honra nas redes:** pena **triplicada** quando cometidos em rede social.
- **Por trás de tudo:** dados vazados vendidos em painéis ilegais e facções com “escritórios” de golpes.
- **Crime de falso advogado.**

O estudo de novos crimes e formas de combate a criminalidade se tornou necessário, afinal nosso Código Penal é de 1940, o que acaba se tornando uma verdadeira colcha de retalhos com a criação de novas normas penais.

2 O QUE SÃO CRIMES CIBERNÉTICOS

Crimes cibernéticos são infrações penais praticadas no ambiente digital, por meio de computadores, redes e dispositivos conectados à internet. As motivações variam: ganho financeiro, espionagem, vandalismo digital e até objetivos políticos (Legale).

A doutrina costuma separá-los em dois grupos:

- **Crimes cibernéticos próprios** — só existem no meio digital, porque o próprio sistema informático é o alvo. Exemplos: invasão de dispositivo, disseminação de malware, ataque de negação de serviço.

- **Crimes cibernéticos impróprios** — delitos tradicionais que usam a tecnologia como meio. Exemplos: estelionato pela internet, ameaça por aplicativo de mensagens, crimes contra a honra em redes sociais.

A distinção tem efeito prático. Nos crimes próprios, o legislador precisou criar tipos penais novos; nos impróprios, bastou, em regra, adaptar os tipos existentes ou criar qualificadoras para o meio digital.

3 PRIMEIROS CRIMES

Os primeiros crimes ligados à informática surgiram nos Estados Unidos, na década de 1960, quando a imprensa e a literatura científica passaram a relatar o uso de computadores em sabotagens e espionagem. Os estudos sistemáticos vieram na década seguinte, e a partir de 1980 as condutas se intensificaram, com manipulação de dados bancários, pirataria de software, abusos em telecomunicações e pornografia infantil (Agência Câmara).

Com a popularização da internet, o crime ganhou escala e alcance transnacional. No Brasil, a resposta legislativa veio aos poucos, quase sempre depois de casos de grande repercussão:

A mudança legislativa não acompanha a tecnologia, esta muda rapidamente, quase que todo dia, enquanto a lei em décadas. A Lei Carolina Dieckmann (2012), e a explosão de golpes por Pix na pandemia, levou a lei 14.155/21, décadas depois dos primeiros casos documentados.

Consequências

Mais da metade dos brasileiros (51%) já foi vítima de fraude digital, segundo relatório da BioCatch divulgado pela CNN Brasil em dezembro de 2025. Só no Pix, os golpes tiraram R$ 4,9 bilhões das vítimas no país.

Os números do mesmo levantamento mostram a velocidade da expansão:

- Golpes relatados no mundo cresceram 65% entre 2024 e 2025.
- Na América Latina, a região mais afetada, o volume se multiplicou por seis em um ano.
- No Brasil, golpes por SMS (*smishing*) aumentaram 14 vezes, e o uso de *deepfakes* cresceu 830%.

- As modalidades mais comuns na região foram o *smishing* e o *vishing* (fraude por ligação telefônica).
- O relatório também aponta a industrialização do crime: há "fábricas de fraude" no Camboja, onde pessoas traficadas são forçadas a aplicar golpes 24 horas por dia, e o modelo já começa a chegar à América do Sul.

As consequências vão além do prejuízo financeiro direto:

- **Para as pessoas:** perda de economias, exposição de dados, dano moral e à reputação, e desconfiança em serviços digitais.
- **Para as empresas:** paralisação de operações, custos de resposta e indenização, sanções da LGPD e perda de credibilidade — como mostra o caso Binance, adiante.
- **Para o Estado:** investigações caras, autoria difícil de provar e criminosos fora da jurisdição nacional.

4 OS CRIMES MAIS COMUNS E SUAS PENAS

A lista abaixo segue os dez crimes apontados pela Legale. As penas foram conferidas com a legislação em vigor; algumas foram aumentadas pela Lei 14.155/2021 e por leis posteriores.

Crime	Base legal	Pena-base
Fraude eletrônica e *phishing*	Art. 171, § 2º-A, CP	Reclusão de 4 a 8 anos e multa
Furto mediante fraude eletrônica	Art. 155, § 4º-B, CP	Reclusão de 4 a 5 anos e multa
Invasão de dispositivo informático	Art. 154-A, CP	Reclusão de 1 a 4 anos e multa
Disseminação de *malware*	Art. 154-A, § 1º, CP	Mesma pena da invasão
Obtenção de dados sigilosos por invasão	Art. 154-A, § 3º, CP	Reclusão de 2 a 5 anos e multa
Estelionato pela internet (forma simples)	Art. 171, CP	Reclusão de 1 a 5 anos e multa

Crime	Base legal	Pena-base
Fraude com ativos virtuais	Art. 171-A, CP	Reclusão de 4 a 8 anos e multa
Extorsão e *ransomware*	Art. 158, CP	Reclusão de 4 a 10 anos e multa
Divulgação de pornografia infantil	Art. 241-A, ECA	Reclusão de 3 a 6 anos e multa
Discurso de ódio em redes sociais	Art. 20, § 2º, Lei 7.716/1989	Reclusão de 2 a 5 anos e multa
Violação de direito autoral com intuito de lucro	Art. 184, § 3º, CP	Reclusão de 2 a 4 anos e multa

Dois pontos merecem atenção. Primeiro, o "roubo de identidade" não tem tipo penal próprio: é enquadrado como fraude eletrônica, estelionato ou falsa identidade (art. 307, CP), conforme o caso. Segundo, a venda de dados pessoais também não tem tipo específico; costuma ser punida pela LGPD na esfera administrativa e, quando há invasão, pelo art. 154-A do Código Penal.

5 INOVAÇÕES JURÍDICAS

O Direito brasileiro saiu de um quadro de quase vazio legal, até 2012, para um sistema que combina tipos penais específicos, proteção de dados e cooperação internacional. As principais inovações:

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. 20º Anuário Brasileiro de Segurança Pública. 2026.

OAB SP. Operação conjunta da OAB SP e Polícia Civil prende organização criminosa que movimentou R$ 10 milhões com Golpe do Falso Advogado. 20/05/2026. Disponível em: https://www.oabsp.org.br/jornaldaadvocacia/26-05-20-1636-operacao-conjunta-da-oab-sp-e-policia-civil-prende-organizacao-criminosa-que-movimentou-r-milhoes-com-golpe-do-falso-advogado

GAZETA DO PARANÁ. Golpe do falso advogado: Polícia Civil faz novo alerta após caso de quase R$ 2 milhões no Paraná. Nov./2026. Disponível em: https://gazetadoparana.com.br/publico/golpe-do-falso-advogado-policia-civil-faz-novo-alerta-apos-caso-de-quase-r-2-milhoes-no-parana

CORREIO BRAZILIENSE. As principais mudanças promovidas pela Lei 15.397/2026 ao Código Penal. Jun./2026. Disponível em: https://www.correiobraziliense.com.br/direito-e-justica/2026/06/7437332-as-principais-mudancas-promovidas-pela-lei-15-397-2026-ao-codigo-penal.html

1. FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. 20º Anuário Brasileiro de Segurança Pública. 2026.

2. OAB SP. Operação conjunta da OAB SP e Polícia Civil prende organização criminosa que movimentou R$ 10 milhões com Golpe do Falso Advogado. 20/05/2026. Disponível em: https://www.oabsp.org.br/jornaldaadvocacia/26-05-20-1636-operacao-conjunta-da-oab-sp-e-policia-civil-prende-organizacao-criminosa-que-movimentou-r-milhoes-com-golpe-do-falso-advogado

3. GAZETA DO PARANÁ. Golpe do falso advogado: Polícia Civil faz novo alerta após caso de quase R$ 2 milhões no Paraná. Nov./2026. Disponível em: https://gazetadoparana.com.br/publico/golpe-do-falso-advogado-policia-civil-faz-novo-alerta-apos-caso-de-quase-r-2-milhoes-no-parana

4. CORREIO BRAZILIENSE. As principais mudanças promovidas pela Lei 15.397/2026 ao Código Penal. Jun./2026. Disponível em: https://www.correiobraziliense.com.br/direito-e-justica/2026/06/7437332-as-principais-mudancas-promovidas-pela-lei-15-397-2026-ao-codigo-penal.html

5. **Tipificação da invasão (Lei 12.737/2012).** Criou o art. 154-A do Código Penal e puniu, pela primeira vez, o acesso não autorizado a dispositivo alheio, independentemente de dano.

6. **Marco Civil da Internet (Lei 12.965/2014).** Obrigou provedores a guardar registros de conexão e de acesso e definiu quando a Justiça pode requisitá-los — base de quase toda investigação de autoria.

7. **LGPD (Lei 13.709/2018).** Impôs às empresas o dever de proteger dados pessoais e de comunicar incidentes, com sanções aplicadas pela ANPD. A segurança da informação deixou de ser opção e virou obrigação legal.

8. **Lei 14.155/2021.** Transformou a invasão em crime punido com reclusão, criou a fraude eletrônica e o furto mediante fraude eletrônica, com penas de 4 a 8 anos, e fixou a competência pelo domicílio da vítima no estelionato por transferência ou Pix.

9. **Marco legal dos criptoativos (Lei 14.478/2022).** Criou o crime de fraude com ativos virtuais (art. 171-A) e atribuiu ao Banco Central a regulação das prestadoras de serviços de ativos virtuais.

10. **Convenção de Budapeste (Decreto 11.491/2023).** Integrou o Brasil à rede internacional de cooperação contra o crime cibernético, o que agiliza a preservação de provas e pedidos a outros países.

11. Lei 15.397/2026 (em vigor desde 04/05/2026). Elevou a pena do estelionato simples, ampliou as hipóteses da fraude eletrônica (redes sociais, telefone, e-mail, duplicação de dispositivo), elevou o furto mediante fraude eletrônica para 4 a 10 anos, criou o crime de cessão de conta laranja e tornou o estelionato crime de ação penal pública incondicionada, dispensando a representação da vítima, que era causa frequente de arquivamento.

12. Responsabilidade das plataformas (STF, Temas 533 e 987, 2025). O STF declarou o art. 19 do Marco Civil parcialmente inconstitucional: em regra, a plataforma passa a responder se, notificada, não remover conteúdo criminoso, mantida a exigência de ordem judicial para crimes contra a honra. Perfis falsos de advogados se enquadram nessa lógica.

Na prática, surgiram também mecanismos fora do Código Penal. O Mecanismo Especial de Devolução (MED) do Banco Central permite bloquear e devolver valores enviados por Pix em caso de fraude. E o caso Binance mostrou outra via: o congelamento de *stablecoins* pela própria empresa emissora, possível porque esses ativos têm um responsável central.

Os desafios continuam. A inteligência artificial barateia golpes com *deepfakes* e ajuda a encontrar falhas de segurança. Criptoativos descentralizados podem ser rastreados, mas não bloqueados. E boa parte dos autores age a partir de países onde a lei brasileira não alcança.

6 O ATAQUE À BINANCE (OUTUBRO DE 2026)

Na quinta-feira, 29 de outubro de 2026, invasores desviaram US$ 6,3 bilhões (cerca de R$ 34 bilhões) da Binance, a maior corretora de criptomoedas do mundo, em aproximadamente 20 minutos (Migalhas). É o maior roubo da história do setor, mais de quatro vezes o de US$ 1,5 bilhão sofrido pela Bybit em fevereiro de 2025 (Acionista).

7 COMO ACONTECEU

Os recursos saíram das carteiras da corretora em pelo menos dez redes diferentes. As empresas de rastreamento TRM Labs e Chainalysis estimaram o total entre US$ 6,36 e 6,4 bilhões (Times Brasil). Entre os ativos levados estavam *stablecoins* (USDT e USDC), bitcoin e ether.

As investigações preliminares indicam que o ataque pode ter explorado uma vulnerabilidade no computador de um engenheiro da própria Binance; outras falhas ainda são apuradas (Migalhas). O FBI atribuiu o roubo à Coreia do Norte (BP Money).

Data	Fato
01/11/2026	Binance retoma os saques, começando por 15 criptoativos (Times Brasil)
31/10/2026	Corretora informa ter recomposto cerca de 90% dos ativos desviados (BP Money)
30/10/2026	Tether congela cerca de US$ 1,57 bilhão em USDT e Circle bloqueia mais de US$ 800 milhões em USDC (Baguete)
29/10/2026	Ataque: US$ 6,3 bilhões saem em cerca de 20 minutos; saques suspensos

Bloqueio e recuperação

Com ajuda de outras empresas do setor, cerca de US$ 3 bilhões foram congelados — o que não equivale a recuperação efetiva. O advogado Flávio Filizzola D'Urso, presidente do Instituto Nacional de Estudos sobre Criptoativos (Inecripto), explica a diferença (Migalhas):

- ***Stablecoins*** como USDT e USDC têm uma empresa emissora, que pode congelar os valores.

- **Criptomoedas descentralizadas** como bitcoin e ether não têm entidade central. A blockchain permite seguir as transações, mas não bloqueá-las.

Para cobrir os clientes, a Binance reservou cerca de US$ 6,34 bilhões, combinando caixa próprio, o fundo de proteção aos usuários (SAFU, de cerca de US$ 1 bilhão) e reservas de seguro (BP Money).

7 LIÇÕES JURÍDICAS

O caso da Binance não se resume ao valor, serve como aviso da vulnerabilidade dos sistemas e sobre o uso crescente da inteligência artificial para encontrar e explorar falhas. A empresa que guarda ativos de terceiros responde pela segurança deles; investir em cibersegurança deixou de ser custo e passou a ser uma obrigação de sobrevivência.

O caso também expõe os limites do Direito Penal. Se os autores atuam a partir de um Estado estrangeiro, a punição depende de cooperação internacional improvável. A resposta efetiva veio de mecanismos privados — congelamento pelas emissoras e fundos de garantia — e não de uma condenação, pouco provável.

8 CRIME DO FALSO ADVOGADO INOVAÇÕES PELA IA

Um golpe já conhecido com novas tecnologias, uma evolução do já conhecido golpe do falso advogado, mas agora, com um nível de sofisticação que tem aumentado o número de vítimas em diferentes estados do país.

Deepfake é uma técnica que permite alterar um vídeo ou foto com ajuda de Inteligência Artificial (IA).

Com ele, por exemplo, o rosto da pessoa que está em cena pode ser trocado pelo de outra; ou aquilo que a pessoa fala pode ser modificado.

Como o golpe funciona

Com uso da IA, criminosos criam perfis falsos nas redes sociais usando nomes, fotos e números, de telefone e de registro reais de advogados.

Produzindo vídeos criados por Inteligência Artificial, com voz e imagem semelhantes às originais.

Os golpistas se aproveitam da fragilidade do cliente, o qual muitas vezes está esperando há anos a solução da sua lide, entram em contato por WhatsApp, usando outro número e agora clonando o número original **caller**

ID spoofing, com a foto do profissional, se utilizam de informações públicas de processos para dar aparência de legitimidade à abordagem, alegando a necessidade de recolher guia para a liberação dos valores depositados em juízo.

Caller ID spoofing (falsificação de identificador de chamadas), é uma técnica utilizada pelos criminosos manipulando de propósito os dados da chamada para que a BINA, do recebedor identifique como sendo de outro número já cadastrado no seu celular, no caso do advogado, normalmente os golpistas conseguem o número do telefone original do profissional através de pesquisas nos processos, ou na ficha de inscrição na OAB.

Em alguns casos os criminosos também costumam enviar links maliciosos que podem direcionar para páginas falsas de bancos ou instalar programas espiões no celular da vítima.

Apesar dos clientes serem avisados do golpe do falso advogado, alguns ainda continuam sendo enganados, principalmente quando o número que aparece na tela é realmente do advogado.

CONSIDERAÇÕES FINAIS

Em seis décadas, o crime cibernético passou de casos isolados de sabotagem a uma indústria capaz de levar US$ 6,3 bilhões em 20 minutos e de atingir metade da população brasileira. O Brasil construiu, desde 2012, uma base legal razoável: tipos penais próprios, penas mais duras, proteção de dados e adesão à Convenção de Budapeste.

Os casos citados mostram, que a lei sozinha não basta. A prevenção técnica, a cooperação internacional e mecanismos ágeis de bloqueio de valores pesam tanto quanto a pena prevista no Código. Para empresas e advogados, a tarefa é dupla: exigir segurança à altura do risco e conhecer as vias — penais, civis e administrativas — para agir rápido quando o ataque acontece.

REFERÊNCIAS

ACIONISTA. Ataque hacker à Binance: veja os impactos para o mercado de criptomoedas. Disponível em: https://acionista.com.br/ataque-hacker-a-binance-veja-os-impactos-para-o-mercado-de-criptomoedas/

AGÊNCIA CÂMARA DE NOTÍCIAS. Conheça a evolução dos crimes cibernéticos. 23/08/2006. Disponível em: https://www.camara.leg.br/noticias/89137-conheca-a-evolucao-dos-crimes-ciberneticos/

BAGUETE. Hackers roubam US$ 6,3 bilhões da Binance. 30/10/2026. Disponível em: https://www.baguete.com.br/noticias/hackers-roubam-us-6-3-bilhoes-da-binance

BP MONEY. Binance repõe 90% dos ativos do hack de US$ 6,3 bi e prevê saques no domingo. 31/10/2026. Disponível em: https://bpmoney.com.br/mercado/cripto/binance-repoe-90-ativos-hack-us-63-bi-saques-domingo/

BP MONEY. Binance retoma saques; FBI atribui ataque à Coreia do Norte. Disponível em: https://bpmoney.com.br/?p=432897

CNN BRASIL. Mais da metade dos brasileiros foram vítimas de golpes digitais em 2024. 05/12/2025. Disponível em: https://www.cnnbrasil.com.br/nacional/brasil/metade-dos-brasileiros-foram-vitimas-de-golpes-digitais-em-2024-diz-estudo/

LEGALE EDUCACIONAL. 10 crimes cibernéticos mais comuns e suas penalidades. Disponível em: https://legale.com.br/blog/10-crimes-ciberneticos-mais-comuns-e-suas-penalidades/

MIGALHAS. Advogado analisa roubo bilionário de criptomoedas da Binance. 01/11/2026. Disponível em: https://www.migalhas.com.br/quentes/467658/advogado-analisa-roubo-bilionario-de-criptomoedas-da-binance

TIMES BRASIL. Ataque à Binance: investigadores tentam recuperar US$ 6,3 bilhões. Disponível em: https://timesbrasil.com.br/colunas/ai-451/binance-roubo-seculo-ia-ataques-frequentes/

TIMES BRASIL. Binance retoma saques após ataque hacker de US$ 6,3 bilhões. 01/11/2026. Disponível em: https://timesbrasil.com.br/cripto-brasil/binance-retoma-saques-apos-ataque-hacker-de-us-63-bilhoes/

ISBN: 978-65-86504-38-5

