

QUANDO O VERDADEIRO PRECISA PROVAR QUE É REAL

O ônus de realidade, o dividendo do mentiroso e a contestação específica da prova digital: uma complementação jurídico-probatória ao Radar Tecnológico nº 6 da ANPD

Jandislei Antonio Genova¹

Resumo

A difusão de conteúdos sintéticos produz externalidade probatória dupla: permite fabricar o falso e depreciar o verdadeiro mediante a alegação de que qualquer registro pode ter sido gerado ou manipulado por inteligência artificial. O artigo investiga como o processo civil brasileiro deve responder quando uma parte impugna prova digital sob a hipótese de *deepfake*. Adota-se pesquisa jurídico-dogmática, qualitativa, documental e interdisciplinar, apoiada no exame integral do Radar Tecnológico nº 6 da Agência Nacional de Proteção de Dados, na legislação processual, em precedentes, padrões de proveniência e literatura especializada. A análise incorpora, como teste de anterioridade e contraste funcional, a minuta de trabalho da Rule 901(c) das Federal Rules of Evidence e o levantamento exploratório com 931 juízes federais norte-americanos, sem tratá-los como direito vigente ou validação do modelo brasileiro. Propõe-se a categoria analítica **ônus de realidade**: custo prático imposto ao proponente de conteúdo genuíno para demonstrar origem, integridade e contexto porque a falsificação sintética se tornou plausível. A contribuição específica consiste em operacionalizar a argumentação exigida pelo art. 436, parágrafo único, do Código de Processo Civil, decompor controvérsias híbridas à luz do art. 429 e associar cinco eixos de especificidade a respostas graduadas. Proveniência criptográfica e detecção algorítmica são tratadas como elementos corroborativos, não certificados de verdade; sua ausência não gera presunção negativa. Conclui-se que a invocação da IA não pode transferir automaticamente o custo probatório e que eventual redistribuição depende de classificação por questão, fundamentação e contraditório.

Palavras-chave: deepfakes; prova digital; ônus de realidade; dividendo do mentiroso; autenticidade; proveniência.

Abstract

WHEN GENUINE EVIDENCE MUST PROVE IT IS REAL: The reality burden, the liar's dividend, and specific challenges to digital evidence

The diffusion of synthetic content creates a dual evidentiary externality: it enables false evidence to be fabricated while depreciating genuine evidence through the claim that any record may have been generated or manipulated by artificial intelligence. This article examines how Brazilian civil procedure should respond when digital evidence is challenged as a possible deepfake. It uses qualitative, documentary, interdisciplinary, and doctrinal legal research based on a full examination of Technological Radar No. 6 issued by the Brazilian National Data Protection Agency, procedural legislation, judicial decisions, provenance standards, and specialized literature. As a prior-art test and functional comparison, the analysis incorporates the working draft of Rule 901(c) of the Federal Rules of Evidence and an exploratory survey of 931 U.S. federal judges, without treating either as binding law or validation of the Brazilian model. The article proposes the analytical category of **reality burden**: the practical cost imposed on the proponent of genuine content to demonstrate origin, integrity, and context because synthetic falsification has become plausible. Its specific contribution is to operationalize the specific grounds required by the sole paragraph of Article 436 of the Brazilian Code of Civil Procedure, decompose hybrid disputes under Article 429, and connect five dimensions of specificity to graduated responses. Cryptographic provenance and algorithmic detection are corroborative elements rather than truth certificates; missing credentials do not support a

¹ Advogado, escritor e pesquisador independente em Direito, inteligência artificial e governança institucional. Autor de *Autonomia decisória na era da IA* (Editora Dialética, 2026). O texto foi desenvolvido pelo autor com apoio editorial e instrumental de inteligência artificial, sob direção, revisão crítica e responsabilidade intelectual humanas.

negative presumption. Invoking AI cannot automatically shift evidentiary costs, and any reallocation requires issue-by-issue classification, reasons, and an opportunity to respond.

Keywords: deepfakes; digital evidence; reality burden; liar's dividend; authenticity; provenance.

1 INTRODUÇÃO

Durante grande parte da história da prova audiovisual, a principal suspeita incidia sobre a falsificação: seria possível cortar, montar, editar ou retirar uma fala de contexto. A inteligência artificial generativa altera a escala, o custo e a qualidade dessa ameaça. Voz, rosto, cenário, documento e assinatura podem ser sintetizados com crescente verossimilhança. O problema, contudo, não termina no conteúdo falso que ingressa no debate público ou no processo. A possibilidade de falsificação retorna sobre o conteúdo autêntico e oferece ao interessado uma defesa de baixo custo: “pode ser IA”.

Essa frase produz uma assimetria. Alegar é barato; demonstrar que um arquivo genuíno permaneceu íntegro pode exigir acesso ao dispositivo de origem, preservação de metadados, obtenção de registros de plataforma, contratação de perito, reconstrução de cadeia de custódia e exposição de dados pessoais de terceiros. Aquele que nega lança uma hipótese; aquele que apresentou o registro pode ser compelido a financiar a restauração de sua credibilidade.

O Radar Tecnológico nº 6 da Agência Nacional de Proteção de Dados (ANPD) oferece um diagnóstico amplo e tecnicamente informado sobre *deepfakes*. O documento reconhece que a maioria dos métodos de detecção baseados em aprendizado profundo ainda enfrenta problemas de confiabilidade e explicabilidade; registra falsos positivos em materiais autênticos e falsos negativos em materiais manipulados; adverte que a área avançou mais na identificação de falsificações do que na demonstração repetível e juridicamente válida de como o detector alcançou sua conclusão; e recomenda abordagem integrada entre tecnologia, análise contextual e supervisão humana (ANPD, 2026, p. 69–75; 141–143).

O Radar também identifica expressamente o impacto da dificuldade de distinguir registros reais de conteúdos sintéticos sobre o Poder Judiciário, as autoridades administrativas e as plataformas, exigindo ferramentas de detecção e critérios jurídicos adaptados (ANPD, 2026, p. 98). Sua contribuição deve ser reconhecida nos limites definidos pela própria série: agregar informações relevantes sem pretensão de esgotar o tema ou firmar posicionamento institucional definitivo (ANPD, 2026, p. 6; 121).

É justamente nesse espaço declarado de não exaustividade que se situa o presente artigo. O Radar examina autenticidade, detecção, rastreabilidade, auditoria e proveniência,

mas não desenvolve as consequências processuais do **dividendo do mentiroso** — o benefício obtido por quem desacredita informação verdadeira invocando a existência de falsificações — nem articula o problema às categorias de falsidade, autenticidade, ônus da prova e impugnação específica previstas no Código de Processo Civil (CPC).² A lacuna é de escopo jurídico-probatório, não um defeito técnico do documento.

No direito brasileiro, a questão não pode ser resolvida pela fórmula genérica segundo a qual “quem apresenta deve provar”. O CPC diferencia a falsidade do documento — formação de documento não verdadeiro ou alteração de documento verdadeiro — da impugnação de sua autenticidade; distribui de modo diverso os respectivos ônus; exige que a arguição de falsidade exponha motivos e meios de prova; e rejeita impugnação genérica (BRASIL, 2015, arts. 427–431 e 436). Uma alegação de *deepfake* pode contestar a autoria atribuída, a origem do arquivo, a integridade do conteúdo, a correspondência com o evento ou todos esses elementos simultaneamente. A etiqueta escolhida pela parte não basta para classificar juridicamente a controvérsia.

Formula-se, portanto, o seguinte problema de pesquisa:

Como impedir que uma alegação genérica de *deepfake* transfira automaticamente ao proponente de prova digital autêntica o custo de demonstrar a realidade, sem inviabilizar a contestação legítima de conteúdos sintéticos?

A hipótese é que a resposta centrada apenas em detectores ou certificados de proveniência é insuficiente. O processo precisa operar em três momentos: a) decompor a impugnação para distinguir falsidade, autenticidade e conteúdo; b) exigir um limiar proporcional de argumentação específica antes de acionar perícia ou deslocar encargos; e c) preservar rotas alternativas de demonstração para pessoas e instituições sem acesso a tecnologias de autenticação.

Propõe-se denominar **ônus de realidade** o custo prático imposto a quem apresenta conteúdo genuíno para demonstrar sua origem, integridade e contexto apenas porque a falsificação sintética se tornou plausível.³ A expressão não designa novo ônus legal e não substitui os arts. 373 ou 429 do CPC. Ela torna visível uma externalidade anterior à decisão

2 O levantamento textual do Radar não localizou as expressões “dividendo do mentiroso”, “ônus da prova”, “ônus probatório”, “argumentação específica”, “art. 429”, “art. 436” ou “Código de Processo Civil”. A constatação delimita o corpus examinado; não implica que a ANPD tivesse dever institucional de desenvolver dogmática processual em publicação voltada principalmente à proteção de dados e à educação tecnológica.

3 A tradução inglesa *reality burden* é funcional e não reivindica termo consolidado. Harris (2024) emprega *burden of proof* para descrever o encargo epistêmico e social gerado por estratégias proativas de verificação. O ônus de realidade aqui proposto é mais estreito: refere-se aos custos incidentes sobre a demonstração processual de conteúdo genuíno e permanece distinto do ônus jurídico previsto em lei.

judicial sobre distribuição do ônus: a passagem da confiança ordinária para a necessidade de reconstrução técnica do verdadeiro.

A tese central é:

A mera possibilidade tecnológica de geração ou manipulação por IA não satisfaz a exigência de argumentação específica do art. 436, parágrafo único, do CPC. A alegação de *deepfake* somente deve produzir medidas probatórias mais gravosas quando individualizar o objeto contestado, apresentar base indiciária técnica, contextual ou custodial, demonstrar nexo com a proposição relevante e formular pedido verificável e proporcional, calibrado pela desigualdade de acesso à evidência.

A contribuição é deliberadamente delimitada. Chesney e Citron (2019) formularam o dividendo do mentiroso; Rini (2020) e Fallis (2021) analisaram a erosão do valor epistêmico dos registros; Harris (2024) demonstrou que soluções proativas de verificação podem transferir aos usuários um pesado encargo e aprofundar desigualdades; Matthews (2024) acrescentou que tais sistemas podem, ao mesmo tempo, reduzir indevidamente a credibilidade de quem não consegue certificar e inflar a de quem possui os meios de fazê-lo. No Brasil, Martinez (2026a) já examinou a prova sintética, o art. 429 do CPC, o dividendo do mentiroso e uma arquitetura de autenticação e perícia. Em maio de 2026, o Comitê Consultivo das Regras de Prova dos Estados Unidos registrou minuta de trabalho da Rule 901(c) com limiar inicial a cargo do impugnante e padrão reforçado de autenticação a cargo do proponente, caso aquele limiar seja satisfeito (UNITED STATES, 2026, p. 11–13). O presente texto não reivindica prioridade sobre o desenho abstrato de dois estágios. Seu avanço específico está em: a) distinguir o ônus epistêmico geral do custo jurídico-processual aqui denominado ônus de realidade; b) reconstruir a solução no regime especial dos arts. 429 e 436 do CPC; c) decompor impugnações híbridas por questões; d) operacionalizar cinco eixos de especificidade e quatro níveis de resposta; e e) vincular medidas tecnológicas a salvaguardas contra uma prova digital de duas classes.

Nas seções seguintes, o artigo delimita método e anterioridade, examina a externalidade probatória dos *deepfakes* e a evidência empírica disponível. Em seguida, reconstrói o regime do CPC, propõe uma matriz de contestação graduada, avalia limites de detecção e proveniência e formula uma arquitetura institucional de preservação do real. Ao final, apresenta cenários de teste, limitações e agenda de validação.

2 MÉTODO, CORPUS E LIMITES

A pesquisa adota método **jurídico-dogmático, qualitativo, documental e interdisciplinar**. Seu objetivo é reconstruir categorias processuais vigentes diante de um

fenômeno tecnológico novo e formular um modelo normativo passível de crítica, teste e aperfeiçoamento. Não se realizou experimento próprio, levantamento sistemático de processos, revisão sistemática ou metanálise. As proposições empíricas são extraídas de pesquisas publicadas e de levantamento institucional exploratório, utilizados para calibrar — não para substituir — a análise jurídica.

O levantamento foi encerrado em 29 de julho de 2026. O corpus foi organizado em seis conjuntos:

1. o texto integral, de 163 páginas, do Radar Tecnológico nº 6 da ANPD, e não apenas o comunicado público de lançamento;
2. o Código de Processo Civil, especialmente os arts. 6º, 139, VI, 357, 369–373, 378, 396–400, 411, 422, 427–431, 434–441 e 464–480;
3. decisões e comunicações oficiais do Superior Tribunal de Justiça sobre integridade, autenticidade, documentação técnica e controle por terceiros de provas digitais, acrescidas de um caso estrangeiro recente utilizado para contraste funcional;
4. o relatório oficial de maio de 2026 do Comitê Consultivo das Regras de Prova dos Estados Unidos, incluindo a minuta da Rule 901(c), as atas de deliberação e os resultados agregados de levantamento do Federal Judicial Center;
5. literatura jurídica, filosófica e empírica sobre *deepfakes*, dividendo do mentiroso, injustiça epistêmica, autenticação e prova;
6. documentos técnicos primários do NIST e da Coalition for Content Provenance and Authenticity (C2PA).

No Radar, realizou-se leitura integral acompanhada de busca textual dirigida por termos em português e inglês relacionados a dividendo do mentiroso, negação, ônus da prova, presunção, autenticidade, falsidade, cadeia de custódia, perícia, proveniência, C2PA, Poder Judiciário e processo. A busca negativa não demonstra, sozinha, inexistência conceitual: por isso, seus resultados foram confrontados com a leitura das seções sobre identificação, limitações técnicas, impactos jurídicos, rastreabilidade e perspectivas. O procedimento oferece uma trilha de auditabilidade, mas não equivale a protocolo bibliométrico.

Na revisão de anterioridade, foram priorizados artigos revisados por pares, legislação, decisões judiciais, documentos institucionais e padrões técnicos. A seleção partiu de referências fundacionais e pesquisa dirigida por combinações de *deepfake evidence*, *liar's dividend*, *burden of proof*, *authenticity challenge*, *synthetic media provenance*, “prova sintética”, “ônus da prova” e “argumentação específica”. A busca incluiu atualização pelo

nome das regras e instituições identificadas nas fontes, o que levou ao relatório oficial de maio de 2026 e à minuta da Rule 901(c). A estratégia é narrativa orientada por problema: não autoriza afirmar que toda a literatura mundial foi exaurida.

As decisões do STJ analisadas provêm do processo penal. Seu uso neste artigo é comparativo e limitado aos requisitos técnicos de integridade, documentação, custódia e controlabilidade por terceiros. Não se transportam automaticamente para o processo civil nem se relativizam a presunção de inocência e o encargo probatório estatal próprios da persecução penal.⁴ Pela mesma razão, o caso *Matter of M.S. (M.H.)*, da Court of Appeals de Nova York, é empregado como caso analítico comparado, sem pretensão de transplante de regras de admissibilidade estrangeiras.

O modelo proposto é submetido a três testes internos:

- **coerência dogmática**, para não confundir falsidade, autenticidade, conteúdo e valor probatório;
- **proporcionalidade procedimental**, para que a contestação legítima seja possível sem premiar alegações oportunistas;
- **justiça distributiva da prova**, para evitar que credenciais tecnológicas se convertam em requisito exclusivo de participação epistêmica.

Os limites são relevantes. A categoria “ônus de realidade” é proposta analítica ainda não validada por dados judiciais brasileiros. O levantamento do Federal Judicial Center informa experiência declarada por juízes federais norte-americanos, não incidência brasileira nem eficácia causal de determinada regra; a baixa frequência observada e o caráter hipotético de parte das respostas impedem extrapolação. Os demais estudos empíricos concentram-se em comunicação política e foram conduzidos fora do Brasil. Padrões de proveniência estão em evolução e sua adoção permanece desigual. A matriz apresentada não é regra legislativa, súmula, protocolo pericial certificado ou instrumento estatisticamente validado. Seu valor inicial é heurístico e dogmático: explicitar decisões que hoje podem permanecer implícitas.

3 ESTADO DA ARTE E DELIMITAÇÃO DA CONTRIBUIÇÃO

3.1 Do falso persuasivo à depreciação do verdadeiro

O debate inicial sobre *deepfakes* concentrou-se no risco de acreditar em eventos inexistentes. Chesney e Citron (2019) demonstraram, porém, que a ameaça possui direção

⁴ Cadeia de custódia possui disciplina expressa no Código de Processo Penal. No processo civil, a expressão é empregada descritivamente para designar documentação da trajetória do arquivo, sem importar automaticamente requisitos, sanções ou efeitos do regime penal.

inversa: a consciência pública de que gravações podem ser fabricadas permite a atores mal-intencionados negar registros verdadeiros e obter um “dividendo do mentiroso”. A falsificação não precisa sequer existir no caso concreto. A tecnologia fornece plausibilidade abstrata à negação.

Rini (2020) situa o problema em uma estrutura epistêmica mais ampla. Registros audiovisuais não servem apenas como prova de eventos particulares; a possibilidade permanente de gravação disciplina práticas testemunhais. Quando o valor dos registros se deteriora, perde-se um “anteparo epistêmico” que sustentava a confiança naquilo que pessoas relatam. Fallis (2021) explica o fenômeno como redução da informação carregada pelo sinal audiovisual: se a probabilidade de existir uma gravação quando o evento não ocorreu aumenta, a gravação distingue menos entre ocorrência e não ocorrência.

Harris (2024) compara duas respostas tecnológicas. A estratégia reativa tenta detectar conteúdo sintético após a criação; a proativa registra e verifica a origem de conteúdos presumidamente autênticos. Para o autor, a primeira enfrenta problema de critério e corrida adaptativa; a segunda pode impor a usuários a responsabilidade de certificar o próprio conteúdo, com efeitos agravados por desigualdades existentes. Matthews (2024) aprofunda o argumento distributivo: a arquitetura proativa pode gerar déficit de credibilidade para quem não certifica e excesso de credibilidade para quem controla meios de verificação. O problema não se reduz à injustiça testemunhal formulada por Fricker (2007, p. 9–29), pois a desvantagem pode decorrer de acesso técnico e não apenas de preconceito identitário; ainda assim, uma infraestrutura que atribui déficit prévio de credibilidade a quem não possui credenciais pode reproduzir seu efeito funcional de exclusão epistêmica.

Esses trabalhos impedem uma solução simplista. Restaurar a confiança não consiste em criar um selo e dividir o mundo entre conteúdo “certificado” e conteúdo “suspeito”. Uma testemunha vulnerável, um cidadão que grava abuso com aparelho antigo, um trabalhador que conserva captura de tela ou uma pequena organização que não adota padrão criptográfico não podem perder valor probatório apenas por estarem fora de uma infraestrutura tecnológica.

3.2 A evolução jurídico-probatória

Capra (2024, p. 2491–2506) descreve como o tema alcançou o Comitê Consultivo das Regras de Prova dos Estados Unidos, diante da dúvida sobre a suficiência dos requisitos ordinários de autenticação. O autor conclui que qualquer padrão reforçado deveria ser acionado somente depois de alguma demonstração pelo impugnante; a afirmação de que o registro “pode ser um *deepfake*” deveria ser juridicamente inócua. A experiência estrangeira

mostra que o problema já não pertence apenas à ética da informação: ele interfere nos limiares de admissibilidade, na decisão de realizar audiência de autenticidade e na relação entre juiz, perito e julgador do fato. O modelo norte-americano, contudo, não pode ser importado sem mediação, pois o CPC brasileiro contém distribuição específica do ônus e proibição expressa de impugnação genérica.

Em 2026, a controvérsia deixou de ser apenas prospectiva. Em *Matter of M.S. (M.H.)*, a Court of Appeals de Nova York, por maioria, considerou insuficientemente autenticados vídeos cuja obtenção envolvia extração por terceiro não ouvido, seleção de trechos de fluxo não recuperado, intervalo de aproximadamente dois anos e meio até a apreensão e ausência de exame técnico qualificado. A decisão não assentou que toda alegação de *deepfake* bastaria: apoiou-se em fragilidades específicas de origem, custódia e fundamento testemunhal. As dissidências, porém, advertiram que a fundamentação poderia estimular invocações de fabricação sintética sem suporte e obrigar o proponente — inclusive vítimas e órgãos com recursos limitados — a contratar especialistas caros para refutá-las (NEW YORK, 2026). A divisão judicial materializa os dois erros que este artigo procura evitar: admitir conteúdo tecnicamente mal fundamentado e converter a possibilidade abstrata de falsificação em encargo universal de autenticação.

Martinez (2026a) realizou contribuição direta ao processo brasileiro. O autor demonstra que a geração de documentos sintéticos desloca a fraude da adulteração posterior para a falsidade originária e propõe arquitetura em duas camadas: autenticidade na origem e perícia multimodal explicável. No plano procedimental, defende incidente de falsidade sintética, distribuição dinâmica do ônus diante de suspeita plausível, padronização de formatos e preservação de metadados. Em trabalho posterior, desenvolve uma arquitetura de remediação em três pilares — ancoragem de proveniência, análise forense e salvaguardas processuais — com foco em documentos financeiros e semiótica jurídica (MARTINEZ, 2026b).

O presente artigo converge com essa agenda, mas enfrenta pergunta residual. “Suspeita plausível” e “argumentação específica” precisam de conteúdo operativo para que o juiz saiba diferenciar uma dúvida tecnicamente articulada de uma negação estratégica. Além disso, a alegação de *deepfake* pode conter simultaneamente impugnação de autoria e arguição de alteração, sujeitas a regras distintas do art. 429. Sem decomposição, a própria escolha vocabular da parte pode definir quem suportará o custo da prova.

3.3 A minuta da Rule 901(c): convergência independente e limites do paralelo

O relatório oficial de maio de 2026 documenta uma minuta de trabalho da Rule 901(c), ainda sem força normativa e não submetida a consulta pública. O desenho possui dois passos. Primeiro, quem sustenta fabricação por IA generativa deve apresentar evidência suficiente para autorizar uma conclusão possível de fabricação e, assim, justificar a abertura da investigação judicial. Se esse limiar for alcançado, o proponente deverá demonstrar ao juiz que o item é, mais provavelmente do que não, autêntico. A minuta também alcança itens oferecidos pelas Rules 901 e 902, prevê aviso prévio razoável e define IA generativa (UNITED STATES, 2026, p. 11–13).

A convergência é relevante, mas não autoriza apresentar a minuta como regra promulgada nem como validação do modelo brasileiro. O próprio Comitê registrou desacordo sobre necessidade, alcance, terminologia, assimetria entre acusação e defesa e risco de legislar antes da formação de casuística suficiente. Em vez de aprovar sua publicação, decidiu manter a matéria em agenda e realizar conferência técnica e prática no segundo semestre de 2026 (UNITED STATES, 2026, anexo, p. 12–13).

O dado judicial disponível reforça simultaneamente a pertinência e a cautela. O Federal Judicial Center consultou o universo de juízes federais norte-americanos de primeira instância, obtendo 931 respostas, equivalentes a 45% dos destinatários. Somente 15 respondentes declararam ter enfrentado contestação de *deepfake*. Entre os que não tinham essa experiência, 82% afirmaram que exigiriam alguma demonstração inicial antes de examinar a alegação de fabricação (UNITED STATES, 2026, anexo, p. 8–9). A frequência declarada é baixa e a maior parte das respostas é contrafactual; o levantamento não mede custos, resultados ou acurácia e não testa a matriz deste artigo. Ele oferece, contudo, evidência institucional de que juízes reconhecem a necessidade de um filtro e de que a matéria ainda carece de base empírica robusta.

Quadro 1 – Convergências e diferenças entre a minuta da Rule 901(c) e o modelo proposto

Elemento	Minuta de trabalho da Rule 901(c)	Modelo proposto para o CPC brasileiro
Gatilho inicial	Evidência suficiente para sustentar uma conclusão possível de fabricação por IA generativa.	Argumentação específica avaliada em cinco eixos, calibrada por controle da evidência, vulnerabilidade e risco.
Efeito do gatilho	O juiz passa a exigir do proponente demonstração de autenticidade pelo padrão “mais provável que não”.	Ativação gradual de preservação, exibição, esclarecimento ou perícia; nenhum deslocamento uniforme e automático do ônus.
Regime de distribuição	Elevação do padrão de autenticação no âmbito das Rules 901 e 902.	Classificação por questão sob os arts. 427–431 e 436; art. 373, § 1º, somente em caráter excepcional e fundamentado.

Elemento	Minuta de trabalho da Rule 901(c)	Modelo proposto para o CPC brasileiro
Escopo material	Alegação de fabricação total ou parcial por IA generativa.	Autoria, origem, integridade, formação sintética, contexto e valor probatório, ainda que a impugnação seja híbrida.
Salvaguarda distributiva	Aviso prévio e oportunidade de resposta; debate ainda aberto sobre assimetrias.	Não exclusividade tecnológica, não negatividade da ausência de credencial e redução da exigência técnica diante de assimetria informacional.
Situação normativa	Minuta mantida em estudo; não é direito vigente.	Proposta doutrinária e heurística; não é regra legislativa, precedente vinculante ou protocolo validado.

Fonte: elaboração do autor (2026), a partir de United States (2026, p. 11–13; anexo, p. 8–13).

A semelhança de estrutura enfraquece qualquer reivindicação de novidade quanto à ideia abstrata de um filtro seguido de resposta reforçada. Ao mesmo tempo, delimita com maior precisão a contribuição: o presente modelo não importa o padrão norte-americano nem cria presunção uniforme; reconstrói um limiar já positivado no art. 436, preserva a alocação específica do art. 429 e organiza medidas anteriores à eventual incidência do art. 373, § 1º.

3.4 Quadro de diferenciação

Quadro 2 – Estado da arte e problema residual

Fonte ou eixo	Contribuição já estabelecida	Questão residual enfrentada neste artigo
Chesney e Citron; Rini; Fallis	Dividendo do mentiroso, erosão do valor informacional e perda do anteparo audiovisual.	Conversão do diagnóstico epistêmico em critério processual brasileiro.
Harris; Matthews	Encargo social da verificação proativa, déficit e excesso de credibilidade, risco de aprofundamento das desigualdades.	Distinção entre encargo epistêmico geral e custo jurídico-processual do proponente de prova genuína.
Capra; <i>Matter of M.S.</i>	Debate normativo e controvérsia judicial sobre autenticação, perícia e risco de uma defesa genérica de <i>deepfake</i> .	Reconstrução a partir da arquitetura específica do CPC brasileiro.
Comitê Consultivo dos EUA (2026)	Minuta de dois estágios da Rule 901(c), pesquisa exploratória com juízes e decisão de aprofundar o estudo antes da publicação.	Cinco eixos de especificidade, quatro níveis de resposta, decomposição por questões e salvaguardas distributivas sem transposição do padrão estrangeiro.
Martinez	Prova sintética, dividendo do mentiroso, art. 429, arquitetura de proveniência e perícia, suspeita plausível.	Conteúdo operacional do art. 436, decomposição de impugnações híbridas e efeitos graduados.
Radar n° 6 da ANPD	Diagnóstico técnico e regulatório; limites de detecção; rastreabilidade, C2PA, auditoria e supervisão humana.	Distribuição do custo probatório, impugnação específica e salvaguardas contra uma prova de duas classes.

Fonte: elaboração do autor (2026).

A originalidade reivindicada é, assim, **incremental, dogmática e operacional**. Não se propõe novo regime geral da prova, não se declara superado o art. 429, não se reivindica

autoria da estrutura abstrata de dois estágios e não se reduz autenticidade a protocolo tecnológico. O artigo procura preencher, no direito processual civil brasileiro, o intervalo entre “é possível contestar” e “a contestação produz consequências”: qual alegação é suficiente, que questão ela realmente suscita, que medida deve ser adotada, qual dispositivo a fundamenta e quem controla os meios de esclarecê-la.

4 A DUPLA EXTERNALIDADE PROBATÓRIA

4.1 Falsificação e depreciação

Os *deepfakes* geram duas externalidades conectadas. A primeira é **positiva para o falsificador**: ferramentas acessíveis reduzem custo, habilidade e tempo necessários para produzir conteúdo que simula sinais de autenticidade. A segunda é **negativa para o titular do verdadeiro**: a difusão social dessas ferramentas diminui o valor inicial de registros genuínos e aumenta o custo de sua aceitação.

A primeira externalidade produz **falso positivo epistêmico**: aceita-se como real o que é sintético. A segunda produz **falso negativo epistêmico**: rejeita-se ou suspende-se indevidamente a confiança no que ocorreu. Uma política exclusivamente voltada à remoção do falso trata apenas metade do problema. O verdadeiro pode perder eficácia antes que qualquer detector seja acionado.

O fenômeno é assimétrico porque a alegação não precisa demonstrar um método específico de geração. A parte pode invocar a evolução geral da tecnologia. Se o juízo responder a toda hipótese abstrata com perícia completa, o sistema recompensa a negação: o impugnante externaliza os custos de sua dúvida ao proponente, ao Judiciário e ao perito.

Define-se **ônus de realidade** como:

o custo informacional, técnico, temporal, financeiro e processual suportado por quem apresenta conteúdo genuíno para demonstrar origem, integridade, autoria e contexto em razão da plausibilidade social ou processual de falsificação sintética.

O conceito possui quatro propriedades. Primeiro, é **de fato**, não *de jure*: pode existir antes de qualquer decisão formal sobre o ônus da prova. Segundo, é **relacional**: cresce conforme o custo da negação diminui e a assimetria de acesso aos registros aumenta. Terceiro, é **distributivo**: recai com maior intensidade sobre quem não dispõe de dispositivos atualizados, assinaturas, repositórios, peritos ou poder de requisição. Quarto, é **governável**: regras de impugnação, preservação, exibição, assistência pericial e interoperabilidade podem reduzi-lo sem declarar toda prova previamente autêntica.

4.2 O que a evidência empírica permite afirmar

Vaccari e Chadwick (2020) realizaram experimento *on-line* com amostra representativa de 2.005 adultos do Reino Unido. Apenas 16% foram classificados como enganados pelo vídeo analisado, enquanto 33,2% expressaram incerteza. Os dois recortes enganosos produziram incerteza em 35,1% e 36,9% dos participantes, contra 27,5% no vídeo completo com revelação educativa. O efeito direto sobre confiança em notícias nas redes sociais não foi significativo; houve efeito indireto mediado pelo aumento da incerteza. O estudo sustenta uma formulação mais precisa: *deepfakes* podem ser mais eficazes em semear indeterminação do que em produzir crença falsa imediata.

Schiff, Schiff e Bueno (2025) conduziram cinco experimentos pré-registrados com mais de 15 mil adultos dos Estados Unidos. Alegações falsas de que escândalos reais seriam desinformação elevaram apoio político em diferentes subgrupos partidários. O efeito foi consistente para relatos textuais, mas amplamente ineficaz diante de vídeos; tampouco houve redução geral da confiança na mídia. Em apenas uma das quatro condições audiovisuais houve efeito positivo relevante, associado à mobilização oposicional. Os resultados confirmam a existência do dividendo do mentiroso, porém rejeitam a extrapolação de que qualquer alegação de *deepfake* destruirá automaticamente a força de todo vídeo.

Essa nuance é essencial. O artigo não sustenta uma “morte da prova audiovisual”. Os estudos tratam de cenários políticos, populações estrangeiras e respostas atitudinais, não de comportamento de magistrados brasileiros. Demonstram plausibilidade e condições de variação. O direito deve antecipar a estratégia sem presumir sua eficácia universal.

4.3 O risco de institucionalizar a dúvida barata

O processo pode amplificar o dividendo do mentiroso mesmo quando o público não acredita imediatamente na negação. Basta que a alegação gere atraso, custo, exposição ou inviabilidade de prova. Uma gravação pode conservar valor persuasivo e ainda assim tornar-se impraticável se sua defesa exigir perícia superior ao valor econômico da causa. O dividendo é, então, procedimental: não deriva da vitória narrativa, mas do preço imposto para que o verdadeiro permaneça utilizável.

Há três formas de institucionalização:

1. **automática**, quando qualquer menção a IA aciona perícia ou deslocamento de encargo;
2. **tecnocrática**, quando apenas detectores ou selos proprietários são aceitos como verificação;

3. **distributiva**, quando a ausência de credenciais reduz a priori o valor de conteúdos produzidos fora de ecossistemas certificados.

O art. 436, parágrafo único, oferece uma barreira normativa já existente contra a primeira forma. A distinção entre proveniência e verdade limita a segunda. A vedação de presunção negativa pela ausência de credenciais, combinada com acesso público a meios de preservação e perícia, enfrenta a terceira.

5 O REGIME PROCESSUAL BRASILEIRO

5.1 Autenticidade, integridade, veracidade e força probatória

A discussão exige separar conceitos frequentemente sobrepostos. **Autoria** identifica a pessoa ou entidade à qual se atribui a produção ou declaração. **Autenticidade**, no sentido estrito adotado pelo CPC, relaciona o documento à autoria que lhe é atribuída; no vocabulário técnico, a expressão muitas vezes alcança também sua origem. **Integridade** indica que o conteúdo relevante permaneceu sem alteração indevida desde determinado marco. **Proveniência** descreve fatos sobre origem, histórico e transformações. **Veracidade** refere-se à correspondência entre a proposição ou representação e o mundo. **Força probatória** é o valor que o julgador atribui ao conjunto após contraditório.

Um arquivo pode ser autenticamente proveniente de determinada organização e conter declaração falsa. Pode estar íntegro desde o envio e ter sido sinteticamente fabricado antes dele. Pode representar evento real, mas omitir contexto decisivo. Pode não possuir assinatura criptográfica e ser corroborado por testemunhas, registros independentes e exame do aparelho. Nenhum indicador isolado responde a todas as perguntas.

O CPC considera autêntico o documento particular quando a firma é reconhecida, quando a autoria é identificada por meio legal de certificação, inclusive eletrônico, ou quando não há impugnação (art. 411). Reproduções fotográficas, cinematográficas, fonográficas ou de outra espécie possuem aptidão probatória se sua conformidade com o original não for impugnada; fotografias digitais e imagens extraídas da internet, se contestadas, demandam autenticação eletrônica ou, quando impossível, perícia (art. 422). Documentos eletrônicos são admitidos e valorados conforme autenticidade, acesso ao teor e legislação de produção e conservação (arts. 439–441).

Essas normas não devem ser lidas isoladamente do art. 436. Se toda afirmação “isso pode ser IA” fosse suficiente para constituir impugnação válida, a proibição de alegação genérica perderia função. A interpretação sistemática é: a consequência técnica prevista no

art. 422, § 1º, pressupõe impugnação processualmente idônea; a idoneidade, nas controvérsias de autenticidade ou falsidade, requer argumentação específica.

5.2 Falsidade e impugnação de autenticidade

O art. 427 define falsidade documental em duas modalidades: formar documento não verdadeiro ou alterar documento verdadeiro. O art. 428 prevê que o documento particular perde sua fé quando a autenticidade é impugnada, enquanto não se comprovar a veracidade. O art. 429 distribui os encargos de modo diferente:

- tratando-se de falsidade ou preenchimento abusivo, o ônus incumbe à parte que a argui;
- tratando-se de impugnação da autenticidade, o ônus incumbe à parte que produziu o documento.

A alegação de *deepfake* pode atravessar as duas hipóteses. Considere um vídeo atribuído ao diretor de uma empresa. A defesa pode: a) negar que a voz e a imagem sejam daquela pessoa, contestando a atribuição; b) admitir que o arquivo se originou do evento ou dispositivo indicado, mas sustentar que foi posteriormente alterado; c) afirmar que o arquivo inteiro foi sinteticamente formado; d) aceitar o arquivo e discutir apenas o sentido, o recorte ou o contexto da fala. Cada proposição possui objeto e consequência próprios.

Não é adequado deixar que a expressão retórica da impugnação defina o regime. “Não sou eu; é IA” pode conter impugnação de autenticidade quanto à atribuição pessoal e arguição de falsidade quanto à formação do arquivo. O juiz deve decompor a controvérsia por questões:

1. **origem material** — de qual dispositivo, sistema, pessoa ou repositório provém o arquivo?
2. **atribuição pessoal ou institucional** — a voz, imagem, assinatura ou declaração pertence a quem se afirma?
3. **integridade temporal** — houve alteração entre captura, armazenamento, transmissão e apresentação?
4. **formação sintética** — o conteúdo foi fabricado ou modificado por método generativo?
5. **contexto e significado** — o registro representa de modo suficiente o fato juridicamente relevante?

Essa decomposição evita duas distorções. A primeira é converter toda falsidade em impugnação de autenticidade para deslocar o ônus ao produtor. A segunda é classificar toda

negação de autoria como falsidade para conservar o encargo com o impugnante. O núcleo factual, não o rótulo, deve orientar o art. 429.

Quadro 3 – Decomposição da alegação de deepfake

Núcleo da controvérsia	Classificação processual predominante	Consequência inicial
“O arquivo não provém da pessoa ou fonte indicada.”	Impugnação de autenticidade ou autoria.	Aplica-se, em princípio, o art. 429, II, após impugnação específica.
“O arquivo provém da fonte indicada, mas foi alterado depois.”	Falsidade por alteração de documento verdadeiro.	Aplica-se, em princípio, o art. 429, I; o impugnante deve expor motivos e meios.
“O arquivo foi integralmente formado sem o evento representado.”	Falsidade por formação de documento não verdadeiro, sem prejuízo da questão de atribuição.	Decomposição entre formação sintética e autoria; encargos por questão.
“A gravação é real, mas o recorte ou contexto distorce seu sentido.”	Impugnação de conteúdo e valoração, não necessariamente autenticidade ou falsidade.	Contraditório, completude contextual e apreciação conjunta.
“Pode ser IA”, sem individualização.	Alegação genérica.	Insuficiente para acionar, por si só, os efeitos dos arts. 422, § 1º, e 429.

Fonte: elaboração do autor (2026), a partir dos arts. 427–429 e 436 do CPC.

“Predominante” é termo deliberado. Impugnações complexas podem acumular questões, e o magistrado deve explicitar a qual fato cada encargo se refere. Distribuição por questão evita a alternativa grosseira de impor a uma única parte o dever de demonstrar toda a história técnica do arquivo.

5.3 O art. 436 como regra antiopportunismo

O art. 436 autoriza a parte intimada a falar sobre documento a impugnar sua admissibilidade, autenticidade, falsidade ou conteúdo. O parágrafo único determina que, nas hipóteses de autenticidade e falsidade, a impugnação se baseie em **argumentação específica**, não se admitindo alegação genérica de falsidade. O art. 431, por sua vez, exige que a arguição de falsidade exponha os motivos e os meios com que se provará o alegado.

Essas regras não exigem que a parte já possua laudo conclusivo. Se exigissem, a impugnação seria inútil justamente para quem precisa do processo para acessar o original, os metadados ou os registros de plataforma. Especificidade também não significa identificar a arquitetura exata, o modelo generativo, a versão do software ou o autor da manipulação. O que se exige é que a dúvida saia do plano da possibilidade universal e ingresse no plano do objeto controvertido.

A mera evolução da IA não é indício particular. Se a possibilidade abstrata bastasse, todos os arquivos seriam permanentemente impugnáveis pelo mesmo fundamento e o parágrafo único do art. 436 seria esvaziado. A regra funciona como mecanismo antiportunismo: não elimina a dúvida, mas exige que aquele que pretende produzir efeitos processuais indique por que **este** conteúdo, **neste** caso, merece verificação adicional.

5.4 O art. 373, § 1º, e a assimetria de acesso

O art. 373, § 1º, admite distribuição diversa do ônus da prova diante de impossibilidade ou excessiva dificuldade de cumprir o encargo ordinário, ou de maior facilidade de obtenção da prova do fato contrário, mediante decisão fundamentada e oportunidade de cumprimento. A dinamização é excepcional, deve incidir sobre questão fática determinada e não pode criar encargo impossível ou excessivamente difícil, conforme o § 2º. Sua função constitucional de equalizar a aptidão probatória não a transforma em cláusula geral para corrigir, sem justificação, toda dificuldade de prova (DORI; CAMBI, 2018, p. 57–66; DIDIER JR.; BRAGA; OLIVEIRA, 2026).

Adota-se aqui uma posição dogmática mínima. O art. 429 contém regra legal especial para falsidade e impugnação de autenticidade e constitui o ponto de partida da alocação do risco. O art. 373, § 1º, não deve ser presumido absolutamente incompatível com essa disciplina, mas tampouco pode apagá-la de modo tácito. Eventual redistribuição exige que a decisão: a) identifique a questão factual específica; b) exponha por que a regra do art. 429 produz impossibilidade, dificuldade excessiva ou descompasso objetivo de acesso; c) demonstre a compatibilidade da medida com o regime especial; d) indique o encargo transferido e os meios razoavelmente disponíveis; e e) seja proferida em tempo de permitir contraditório e efetiva desincumbência.

Essa distinção separa **ônus de provar**, **dever de cooperar** e **poder instrutório**. A parte ou o terceiro que controla arquivo, dispositivo ou log pode ser obrigado a preservá-lo, exibi-lo ou esclarecer sua trajetória sem que isso, por si só, altere a regra de julgamento. Antes de cogitar redistribuição, o juízo pode:

- exigir especificação da impugnação, nos termos dos arts. 431 e 436;
- delimitar as questões de fato e os meios de prova no saneamento, conforme o art. 357, II e IV;
- ordenar preservação e exibição de original, dispositivo, registros ou documentação, observados sigilo, minimização e proteção de dados, com apoio nos arts. 369, 370 e 396–400;

- exigir cooperação de quem controla a evidência, à luz dos arts. 6º e 378, e ajustar a produção probatória pelo art. 139, VI;
- abrir contraditório documental dirigido, conforme os arts. 434–437;
- determinar perícia proporcional, delimitar objeto e formular quesitos verificáveis, nos termos dos arts. 464–480;
- valorar recusa injustificada, perda ou não exibição segundo o regime aplicável, sem converter automaticamente a lacuna em prova de falsidade.

Somente depois de classificar a controvérsia e avaliar medidas menos disruptivas, se persistirem os pressupostos legais, a distribuição dinâmica poderá ser debatida com fundamentação autônoma. Nem vulnerabilidade econômica isolada nem mera alegação de que a outra parte “tem mais tecnologia” bastam. Devem existir relação concreta entre controle informacional e fato controvertido, viabilidade real de produção pelo novo onerado e ausência de surpresa. Essa cautela impede que a categoria “ônus de realidade” seja convertida em inversão automática contrária à lei.

5.5 A jurisprudência sobre integridade e controlabilidade

Embora não trate especificamente de *deepfakes*, a jurisprudência recente do STJ oferece parâmetros úteis. No AgRg no HC 828.054/RN, a Quinta Turma considerou inadmissíveis dados de telefone celular obtidos sem metodologia e documentação capazes de assegurar integridade. O tribunal exigiu registro das fases de reconhecimento, coleta, acondicionamento, transporte e processamento, com identificação dos responsáveis, métodos e ferramentas (BRASIL, 2024).

No AgRg no HC 1.014.212/ES, a Sexta Turma entendeu necessária perícia quando havia dúvida razoável sobre integridade e autenticidade de capturas de tela e imagens de videomonitoramento centrais à imputação. A confiabilidade não decorreria da autoridade de quem acessou o conteúdo, mas da possibilidade de controle técnico por terceiros; autorização judicial e identificação do agente não supririam a ausência de documentação técnica (BRASIL, 2026).

Os precedentes mostram duas ideias relevantes: **autoridade não substitui verificabilidade** e **integridade depende de processo documentado**. Mas também revelam limite: não autorizam concluir que qualquer alegação abstrata imponha perícia, nem transportam ao processo civil as consequências próprias da prova penal. No próprio AgRg no HC 1.014.212/ES, o relator distinguiu imagens extraídas diretamente de sistema de gravação e identificadas quanto à origem, cuja contestação de cortes ou lacunas não exigiria

necessariamente a mesma perícia. O contexto probatório, o controle da fonte e o tipo de dúvida importam.

6 O LIMAR DE ARGUMENTAÇÃO ESPECÍFICA

6.1 Cinco eixos de especificidade

Para operacionalizar o art. 436 sem criar formalismo impossível, propõe-se avaliar a impugnação por cinco eixos. Eles não formam lista rígida e cumulativa; funcionam como estrutura de justificação calibrada pela informação acessível à parte.

a) Individualização do objeto

A parte deve indicar o arquivo, página, trecho, quadro, intervalo temporal, voz, assinatura, imagem ou metadado contestado. Em conteúdo longo ou conjunto documental extenso, a expressão “tudo é falso” não permite resposta útil, salvo quando a própria tese seja a fabricação integral e haja base concreta para tanto.

b) Hipótese controvertida

É necessário dizer se a dúvida recai sobre autoria, origem, alteração posterior, geração integral, sincronização de voz e imagem, inserção de elemento, corte, metadado ou contexto. Não se exige nomear o modelo de IA. Exige-se tornar inteligível o mecanismo alegado.

c) Base indiciária articulável

A base pode ser técnica, custodial ou contextual. São exemplos: incompatibilidade de metadados; ausência ou ruptura inexplicada da cadeia de transmissão; conflito entre versão e sistema; artefato de áudio ou imagem; divergência de iluminação ou sincronização; inexistência do evento no local indicado; declaração incompatível com registros independentes; falha de assinatura; inconsistência entre original e cópia; notícia de comprometimento do repositório. A simples existência social de *deepfakes* não constitui indício individual.

d) Nexa com a proposição relevante

A irregularidade apontada deve afetar o fato que o documento pretende provar. Compressão, recodificação, corte ou ausência de metadados podem decorrer de plataformas e não demonstram, isoladamente, fabricação do conteúdo. A impugnação deve explicar por que o sinal compromete autoria, integridade, formação ou contexto relevante.

e) Pedido de verificação proporcional e controle da evidência

Quando possível, a parte deve indicar o que pretende verificar: exibição do arquivo nativo; preservação de logs; comparação com dispositivo; esclarecimento de origem; perícia em intervalo específico; requisição a plataforma; oitiva de testemunha; confronto com registros independentes. O juízo deve considerar quem controla esses elementos. Quanto maior a assimetria de acesso, menor pode ser a exigência de detalhe técnico do impugnante e maior o dever de cooperação do custodiante.

A matriz preserva duas proposições em tensão. **Especificidade não é prova antecipada da falsidade. E possibilidade tecnológica não é especificidade.**

6.2 Graduação e efeitos processuais

Propõe-se uma escala de quatro níveis. Seu objetivo não é transformar decisão judicial em pontuação automática, mas associar a intensidade da dúvida a medidas proporcionais.

Quadro 4 – Matriz graduada de contestação de conteúdo digital

Nível	Características	Resposta processual sugerida	Fundamento jurídico predominante
0 — Alegação genérica	Invoca IA ou manipulação possível sem individualizar objeto, hipótese, indício ou nexo.	Não aciona, por si só, perícia, perda de fé ou deslocamento de encargo; admite convite à especificação e preservação cautelar mínima.	Arts. 431 e 436, parágrafo único; arts. 6º e 139, VI, para adequação procedimental sem redistribuição automática.
1 — Contestação inicial específica	Individualiza objeto e hipótese; apresenta ao menos um indício técnico, custodial ou contextual articulado ao fato.	Contraditório dirigido, preservação, exibição, esclarecimentos e delimitação dos pontos controvertidos.	Arts. 357, 369, 370, 378, 396–400 e 434–437.
2 — Contestação qualificada	Acrescenta conflito verificável de metadados, proveniência, cadeia, versões, registros independentes ou análise técnica preliminar.	Perícia delimitada ou diligência técnica proporcional; somente depois, exame fundamentado de eventual redistribuição por questão.	Arts. 464–480; arts. 429 e 373, §§ 1º e 2º, conforme classificação e pressupostos.
3 — Contestação positivamente sustentada	Apresenta resultado forense reproduzível, incompatibilidade objetiva ou conjunto convergente de evidências.	Exame técnico aprofundado e decisão fundamentada sobre admissibilidade, eficácia ou valor probatório; eventual apuração de fraude, sem sanção automática.	Arts. 427–431, 436 e 464–480; arts. 79–81 apenas se demonstrada conduta consciente e abusiva.

Fonte: elaboração do autor (2026).

O nível 0 não deve ser simplesmente apagado dos autos. O juiz pode convidar à especificação, especialmente quando a parte é vulnerável, não possui assistência técnica ou depende de informação controlada pelo adversário. A consequência rejeitada é a produção

automática de custo grave. Preservar um arquivo enquanto se esclarece a alegação é medida menos onerosa que ordenar imediatamente perícia integral.

O nível 1 exige resposta focalizada. A parte produtora pode apresentar o arquivo nativo, descrever a captura, juntar registro contemporâneo ou indicar testemunhas. O nível 2 torna razoável intervenção técnica, mas a perícia deve responder a quesitos delimitados. O nível 3 não dispensa contraditório, não converte detector em oráculo e não determina exclusão automática; ele eleva a necessidade de exame e exige que a consequência seja vinculada ao regime jurídico da questão efetivamente demonstrada.

6.3 Calibração por vulnerabilidade e risco

Um limiar uniforme seria injusto. A mesma exigência documental não pode recair sobre cidadão que grava violência inesperada e sobre banco que produz documento em infraestrutura controlada; sobre consumidor que recebe áudio por aplicativo e sobre órgão público que gera arquivo oficial; sobre pequena causa e sobre processo capaz de privar liberdade, patrimônio substancial ou direitos fundamentais.

A calibração deve considerar:

- gravidade e reversibilidade da decisão;
- valor e centralidade da prova;
- capacidade econômica e técnica das partes;
- controle sobre dispositivo, original, plataforma e registros;
- existência de dever institucional prévio de documentação;
- risco de destruição ou volatilidade do conteúdo;
- dados pessoais, sigilo, segurança e direitos de terceiros;
- alternativas menos invasivas de confirmação.

Nos contextos de alto impacto, a instituição que produz sistematicamente registros pode ter deveres de preservação mais intensos. Isso não significa que todo cidadão deva antecipar litígio e operar como laboratório forense. A arquitetura deve exigir mais de quem controla infraestrutura e menos de quem apenas presencia o fato.

6.4 Sanções, boa-fé e prudência

Alegações dolosamente falsas podem violar boa-fé processual e configurar litigância de má-fé, conforme o caso. A sanção, porém, não deve ser usada para desencorajar contestações razoáveis em campo tecnicamente incerto. Confundir erro, dúvida e dolo produziria efeito inibidor sobre vítimas reais de manipulação.

O tratamento adequado é escalonado: oportunidade de especificar; dever de preservar; explicitação dos fundamentos; diligência proporcional; e somente diante de comportamento consciente e processualmente abusivo, aplicação das consequências previstas em lei. O processo precisa ser resistente tanto ao conteúdo sintético quanto à acusação sintética de que o conteúdo é falso.

7 DETECÇÃO ALGORÍTMICA: EVIDÊNCIA PROBABILÍSTICA, NÃO VEREDITO

7.1 O diagnóstico do Radar

O Radar nº 6 registra que muitos detectores baseados em aprendizado profundo não são confiáveis nem explicam adequadamente o resultado; mesmo assim, fornecem rótulo, percentual ou pontuação de falsificação (ANPD, 2026, p. 69). A aparência numérica pode exceder a validade real do sistema. Desempenho de laboratório não se transfere automaticamente a arquivos comprimidos, novos métodos de falsificação ou materiais históricos.

Entre as limitações, o documento identifica falsos positivos em conteúdos autênticos, falsos negativos em *deepfakes*, vieses decorrentes de bases homogêneas e variabilidade diante de resolução, duração, etnia, gênero e idade (ANPD, 2026, p. 71–75). A consequência processual é direta: uma pontuação não pode ser interpretada sem informação sobre modelo, versão, conjunto de validação, limiar, taxa de erro, domínio de uso e tratamento do arquivo.

O NIST organiza as respostas técnicas em autenticação e rastreamento de proveniência, rotulagem e marca d'água, detecção, testes, auditoria e manutenção (CHANDRA et al., 2024). A pluralidade confirma que não existe solução única. Detecção é uma camada entre outras e deve ser avaliada no sistema sociotécnico.

7.2 Requisitos mínimos de uso forense

Um detector utilizado como suporte probatório deve permitir, conforme a relevância e o risco:

1. identificação do fornecedor, modelo, versão e data;
2. preservação do arquivo de entrada e de seu hash;
3. registro de transformações anteriores à análise;
4. descrição do domínio, base e métricas de validação pertinentes;
5. taxa de falsos positivos e negativos no cenário comparável;
6. limiar decisório e interpretação da pontuação;
7. repetibilidade ou, quando inviável, documentação suficiente para auditoria;

8. acesso das partes aos resultados, limitações e quesitos;
9. revisão por perito competente e possibilidade de contradita;
10. declaração de conflitos, dependências e cadeia de custódia.

Nenhum requisito transforma o detector em prova conclusiva. Eles tornam sua contribuição controlável. A avaliação humana também não funciona como selo automático: precisa ser competente, independente, documentada e exposta ao contraditório.

7.3 O problema do critério

Harris (2024) denomina “problema da roda” a dificuldade de justificar o critério pelo qual um detector é considerado confiável diante da evolução adversarial. O histórico de acertos pode perder valor quando os geradores mudam e o domínio se desloca. Matthews (2024) contesta parte da força dessa objeção contra estratégias reativas, mas concorda que a solução proativa suscita injustiças de credibilidade.

Para o processo, a conclusão prudente é dupla. Não se deve rejeitar todo detector por falibilidade; perícias tradicionais também lidam com incerteza. Tampouco se deve converter acurácia agregada em certeza sobre o caso singular. A saída é **corroboração multimétodo**, documentação, comparação contextual e explicitação da incerteza.

8 PROVENIÊNCIA: O HISTÓRICO NÃO É A VERDADE

8.1 Potencial dos Content Credentials

O padrão C2PA registra, em estrutura criptograficamente vinculada, afirmações sobre origem, ferramentas, modificações e uso de IA. Assinaturas, hashes e vínculos entre conteúdo e manifesto tornam alterações detectáveis e podem preservar histórico de criação (C2PA, 2025b). O Radar nº 6 o apresenta como mecanismo padronizado de rastreabilidade, apto a favorecer transparência, verificação, auditoria e responsabilização (ANPD, 2026, p. 141–143).

Em contexto probatório, esses registros podem responder perguntas importantes: o arquivo apresentado corresponde ao que foi assinado? Qual ferramenta realizou determinada edição? Houve transformação após o marco indicado? Que entidade assinou as afirmações? O manifesto se vincula a qual ativo?

8.2 Limites reconhecidos pelo próprio padrão

O explicador oficial da C2PA estabelece limites que precisam integrar qualquer uso jurídico. Content Credentials não formulam juízo de valor sobre a verdade dos dados de proveniência; verificam se as informações estão bem formadas, não foram adulteradas, são

válidas e foram assinadas por entidade reconhecida em determinada lista de confiança. O padrão não é solução completa para desinformação e deve complementar letramento, verificação de fatos e análise forense (C2PA, 2025a).

O próprio documento rejeita a criação de ecossistema de mídia em duas classes e afirma que não se deve presumir confiabilidade ou desconfiança apenas pela presença ou ausência de credenciais. Proveniência pode estar incompleta, metadados podem ser removidos, identidade humana não é necessariamente incluída e a história do ativo não revela, sozinha, se o conteúdo é verdadeiro, exato ou factual (C2PA, 2025a).

Essas limitações são juridicamente decisivas:

- assinatura de proveniência demonstra vínculo e integridade de afirmações, não ocorrência do evento;
- o emissor pode estar enganado, comprometido ou ser indigno de confiança;
- edição legítima pode romper ou omitir a cadeia;
- conteúdo sem credenciais pode ser autêntico;
- conteúdo com credenciais pode representar evento encenado, declaração falsa ou contexto enganoso;
- listas de confiança incorporam decisões institucionais sobre quem merece reconhecimento.

8.3 A privatização do lastro probatório

Se poucos fabricantes de dispositivos, plataformas, autoridades certificadoras ou listas privadas definirem quais assinaturas aparecem como confiáveis, a arquitetura técnica passará a organizar a autoridade epistêmica. Não se trata apenas de compatibilidade de software. Quem ingressa na lista ganha capacidade de fornecer um atalho de credibilidade; quem permanece fora pode ser obrigado a provar por meios mais caros.

A literatura sobre plataformização mostra que serviços privados podem adquirir características de infraestrutura ao mesmo tempo que infraestruturas passam a operar segundo lógicas de plataforma, combinando escala, programabilidade e controle concentrado (PLANTIN et al., 2018). Aplicado à prova, esse movimento significa que uma camada privada concebida para circulação de conteúdo pode tornar-se porta quase obrigatória de autenticação sem ter sido desenhada segundo garantias de contraditório, continuidade e acesso universal.

A solução não é rejeitar padrões privados nem estatizá-los integralmente. É impedir dependência exclusiva. Uma arquitetura jurídica adequada requer padrões abertos, interoperabilidade, pluralidade de âncoras de confiança, auditoria, portabilidade, possibilidade de verificação independente e rotas probatórias para conteúdos não credenciados.

Esse ponto conecta proveniência à influência cognitiva estrutural. A interface que exhibe um selo verde, um alerta vermelho ou ausência silenciosa organiza a percepção antes da avaliação consciente. A tecnologia não decide formalmente o mérito, mas define quais conteúdos chegam ao julgador com credibilidade inicial reforçada ou diminuída. A autonomia decisória depende, assim, do domínio das condições materiais pelas quais a escolha e a valoração se tornam possíveis (GENOVA, 2026).

8.4 Regra de não exclusividade e não negatividade

Propõem-se duas salvaguardas:

Regra de não exclusividade: nenhum padrão, detector, lista de confiança ou fornecedor deve constituir via única de demonstração de autenticidade, integridade ou proveniência.

Regra de não negatividade: a ausência de credencial, marca d'água, assinatura ou metadado não pode, isoladamente, gerar presunção de falsidade, manipulação ou menor credibilidade.

As regras não atribuem igual valor a toda prova. Um arquivo com cadeia tecnicamente robusta pode receber maior peso após contraditório. O que se rejeita é a inferência automática de que “não certificado” significa “suspeito”. O peso positivo de uma credencial não deve produzir, por reflexo necessário, punição epistêmica a quem não a possui.

9 UMA ARQUITETURA PÚBLICA DE PRESERVAÇÃO DO REAL

9.1 Princípios de desenho

Responder ao ônus de realidade exige mais que nova regra de admissibilidade. É necessário organizar o ciclo de vida da evidência para que origem, transformações e contexto possam ser demonstrados sem dependência absoluta de um único fornecedor. A arquitetura proposta é **pública** não porque todos os componentes devam ser operados pelo Estado, mas porque suas regras, interfaces, testes e mecanismos de contestação precisam ser governáveis à luz do interesse público.

Sete princípios orientam o desenho:

1. **neutralidade tecnológica** — a norma define propriedades verificáveis, não marca ou produto obrigatório;

2. **interoperabilidade e portabilidade** — registros podem ser exportados, validados e preservados em formatos abertos;
3. **pluralidade de confiança** — diferentes âncoras públicas, privadas, comunitárias e profissionais podem coexistir;
4. **não exclusividade probatória** — credenciais complementam, mas não substituem, testemunhos, contexto, perícia e corroboradores;
5. **proteção de dados por desenho** — minimização, retenção proporcional, divulgação seletiva e segurança limitam vigilância e exposição;
6. **auditabilidade e contestabilidade** — decisões de validação, falhas e listas de confiança podem ser examinadas por terceiros autorizados;
7. **inclusão epistêmica** — pessoas sem infraestrutura certificada conservam rotas acessíveis para demonstrar autenticidade.

9.2 Camadas funcionais

Quadro 5 – Arquitetura institucional de preservação do real

Camada	Função e controles	Salvaguarda contra captura ou exclusão
Captura e origem	Registro opcional de dispositivo, data, local, hash, ferramenta e identidade institucional; assinatura quando disponível.	Minimização de dados; identidade humana seletiva; nenhuma presunção negativa pela ausência.
Preservação e custódia	Arquivo nativo, hash, histórico de transferências, responsáveis, transformações e política de retenção.	Formatos exportáveis; repositórios redundantes; dever proporcional a quem controla infraestrutura.
Corroboração contextual	Testemunhas, registros independentes, documentos contemporâneos, sensores, localização e sequência de eventos.	Evita reduzir verdade a metadado ou selo; permite validar conteúdos legados.
Análise técnica	Detectores múltiplos, exame de sinais, comparação de versões e avaliação humana documentada.	Divulgação de versão, métricas, incerteza e conflitos; proibição de veredito automático.
Porta processual	Matriz de especificidade, classificação pelo art. 429, preservação, exibição, perícia proporcional e contraditório.	Impede que a hipótese abstrata de IA produza custo integral ou perda automática de valor.
Governança da confiança	Critérios públicos para listas, acreditação, auditorias, incidentes, revogação e recursos.	Pluralidade de validadores; revisão independente; nenhuma autoridade privada universal.
Acesso e assistência	Ferramentas abertas, laboratórios públicos ou conveniados, assistência técnica e modelos de protocolo.	Reduz desigualdade entre grandes instituições e cidadãos, pequenas empresas ou grupos vulneráveis.

Fonte: elaboração do autor (2026).

As camadas não precisam existir em todos os casos com a mesma intensidade. Uma fotografia cotidiana pode depender principalmente de contexto, arquivo nativo e testemunho. Registros de órgão público, banco, hospital ou plataforma de grande porte podem justificar trilhas mais robustas desde a origem, pois essas entidades controlam processos recorrentes e produzem evidência em escala.

9.3 Proteção de dados e o paradoxo da autenticação

Autenticar pode aumentar a exposição que se pretendia combater. Registrar identidade, localização, dispositivo e histórico de edição em cada conteúdo cria dados úteis para vigilância, perseguição ou discriminação. Defensores de direitos humanos, denunciante, vítimas de violência e jornalistas podem precisar provar integridade sem revelar publicamente sua identidade.

A arquitetura deve separar:

- **integridade do ativo**, verificável por hash e assinatura;
- **qualidade do emissor**, avaliada pela cadeia de confiança;
- **identidade real**, divulgada somente quando necessária e juridicamente autorizada;
- **contexto probatório**, acessível conforme contraditório, sigilo e minimização.

Credenciais seletivas, pseudônimos verificáveis, divulgação por atributos, custódia por terceiro confiável e revelação apenas ao juízo podem compatibilizar autenticidade e segurança. A proveniência não pode converter todo cidadão em objeto de rastreamento permanente.

9.4 Um protocolo reutilizável para o setor público

A proposta pode ser traduzida em artefato público reutilizável composto por:

1. esquema mínimo de metadados e cadeia de transformação;
2. procedimento de geração e verificação de hashes;
3. formulário de documentação de captura, transferência e edição;
4. matriz de contestação específica;
5. quesitos-modelo para perícia;
6. testes de interoperabilidade e exportação;
7. política de retenção, acesso, incidente e revogação;
8. salvaguardas de privacidade e acessibilidade;
9. conjunto de cenários de conformidade;
10. implementação de referência em código aberto, quando tecnicamente cabível.

Esse artefato não precisaria armazenar o conteúdo probatório nem funcionar como autoridade nacional única. Seu valor estaria em padronizar processos auditáveis e permitir adoção por órgãos distintos, inclusive em bancos de soluções públicas de IA. A ideia central é separar **infraestrutura de confiança** de **monopólio da verdade**.

10 CENÁRIOS DE APLICAÇÃO E TESTE

Uma proposta normativa precisa demonstrar como se comporta em casos difíceis. Os cenários seguintes não são precedentes reais; são construções analíticas destinadas a expor consequências e possíveis erros.

Quadro 6 – Aplicação demonstrativa

Cenário	Risco de aplicação inadequada	Resposta proposta
Cidadã apresenta vídeo de abuso gravado por aparelho antigo, sem credenciais.	Desvalorizar o arquivo por ausência de selo e impor perícia inacessível.	Preservar arquivo nativo; colher contexto e testemunhos; comparar registros independentes; exigir impugnação específica; oferecer assistência técnica se necessária.
Órgão público apresenta gravação oficial, mas não conserva logs ou versão do sistema.	Transferir ao administrado o prejuízo da falha institucional de documentação.	Exigir exibição e explicar a lacuna; considerar dever prévio de custódia; valorar a perda conforme regime aplicável; não presumir integridade pela autoridade da fonte.
Parte junta captura de conteúdo hospedado em plataforma que controla o original.	Obrigar uma das partes a provar fato cujo registro está com terceiro.	Preservação urgente, requisição ou exibição direcionada; registro do URL, data e contexto; perícia somente no ponto controvertido.
Empresa apresenta vídeo com Content Credentials válidos.	Tratar credencial como prova de que o evento representado ocorreu e é verdadeiro.	Reconhecer integridade das afirmações assinadas e examinar emissor, contexto, encenação, completude e registros externos.
Pessoa pública afirma apenas que gravação comprometedora “pode ser deepfake”.	Produzir dividendo procedimental mediante negação de custo quase zero.	Aplicar nível 0; intimar para especificar; não deslocar encargo nem ordenar perícia integral com fundamento exclusivo na possibilidade abstrata.
Vítima recebe áudio suspeito, sem acesso ao aparelho do remetente, e aponta incoerência contextual.	Exigir identificação do modelo generativo ou laudo privado antes de preservar a prova.	Reconhecer assimetria; admitir contestação inicial específica; preservar registros, requisitar dados disponíveis e delimitar exame técnico.

Fonte: elaboração do autor (2026).

Os cenários revelam que a matriz não favorece invariavelmente produtor ou impugnante. Ela favorece **explicitação**. Instituições com controle técnico recebem deveres mais intensos; cidadãos sem infraestrutura preservam vias alternativas; partes que alegam

deepfake de modo concreto obtêm diligências; alegações genéricas não externalizam automaticamente todo o custo.

11 INFLUÊNCIA COGNITIVA E SOBERANIA PROBATÓRIA

11.1 A valoração começa antes da decisão

A autenticidade não é examinada em vazio. Interface, selo, alerta, ordem de apresentação, reputação da fonte e linguagem pericial moldam a percepção do julgador. Um percentual de “92% sintético” ou um ícone de “conteúdo verificado” oferece âncora cognitiva antes da análise das limitações. A autoridade aparente do sistema pode substituir a pergunta pelo método.

Esse mecanismo integra, como extensão normativa da análise e não como resultado empírico medido neste estudo, a **influência cognitiva estrutural**: a arquitetura organiza as condições de atenção, confiança e escolha sem retirar formalmente do sujeito a competência para decidir (GENOVA, 2026). O magistrado continua livre para valorar a prova, mas a infraestrutura pode determinar quais sinais se tornam visíveis, quais incertezas são omitidas e quais alternativas exigem esforço adicional.

Por isso, transparência não se resume a mostrar a existência de IA. É necessário revelar:

- quem definiu a lista de confiança;
- quais versões e métricas sustentam o detector;
- quais dados foram removidos pela plataforma;
- quem conserva o original e os logs;
- qual consequência a interface atribui à ausência de credencial;
- quem pode contestar, auditar e substituir o componente.

Uma decisão formalmente humana pode permanecer materialmente condicionada por uma arquitetura que pré-seleciona sinais de credibilidade. Preservar autonomia exige domínio institucional das condições que tornam a avaliação possível.

11.2 Soberania probatória

Denomina-se **soberania probatória**, em sentido funcional, a capacidade de um sistema jurídico de definir e controlar os critérios pelos quais evidências socialmente

relevantes são preservadas, verificadas, contestadas e valoradas, sem dependência paralisante de fornecedor, plataforma ou lista de confiança estrangeira.⁵

Soberania não exige isolamento tecnológico. Pode utilizar C2PA, ferramentas comerciais, laboratórios privados e cooperação internacional. Exige, porém:

- possibilidade de auditoria e substituição;
- preservação em formatos acessíveis;
- pluralidade de fornecedores e validadores;
- competência pública para formular quesitos e avaliar limitações;
- acesso a registros relevantes sob salvaguardas;
- rotas de contingência se serviço, empresa ou padrão desaparecer;
- capacidade de reconhecer prova autêntica fora do ecossistema dominante.

Quando o Estado não pode auditar o detector, interpretar o manifesto, acessar o log ou migrar o registro, a decisão judicial conserva autoridade formal, mas depende materialmente de critérios definidos fora de seu alcance. A captura da soberania probatória não ocorre apenas quando alguém determina o resultado; ocorre quando a instituição perde a capacidade de reconstruir por que determinada evidência recebeu crédito.

11.3 Responsabilidade pela infraestrutura

O ônus de realidade não deve recair exclusivamente sobre litigantes. Plataformas que removem metadados, organizações que não preservam originais, fabricantes que restringem exportação, fornecedores que ocultam taxas de erro e órgãos que adotam sistemas sem protocolo contribuem para o custo coletivo de verificação.

A responsabilidade institucional adequada é preventiva:

- preservar sem reter excessivamente;
- documentar sem expor indiscriminadamente;
- permitir acesso sem divulgar segredos além do necessário;
- testar sem prometer infalibilidade;
- sinalizar incerteza sem desqualificar automaticamente;
- conservar alternativas para quem está fora da infraestrutura.

Assim, a arquitetura de prova passa a ser compreendida como parte da governança da IA, e não como problema que surge apenas depois do dano.

⁵ “Soberania probatória” é utilizada, sem reivindicação de originalidade terminológica, como categoria funcional derivada da soberania decisória e tecnológica. Não designa competência jurisdicional, regra de direito internacional ou poder de declarar a verdade por autoridade. Refere-se à capacidade institucional de controlar condições de preservação e verificação.

12 LIMITES, FALSIFICABILIDADE E AGENDA EMPÍRICA

12.1 O que refutaria ou reduziria a tese

A categoria proposta perderia utilidade se estudos de processos brasileiros demonstrassem que alegações genéricas de *deepfake* são sistematicamente rejeitadas sem custo adicional, que a autenticação ordinária resolve as controvérsias de modo acessível ou que a existência de conteúdo sintético não altera distribuição, tempo, custo ou valoração da prova. Também seria enfraquecida se a matriz não produzisse decisões mais consistentes que o juízo não estruturado ou se gerasse barreiras desproporcionais a vítimas de falsificação.

A arquitetura técnica precisaria ser revista se:

- a documentação exigida criasse mais risco de privacidade do que ganho probatório;
- padrões abertos não reduzissem dependência;
- credenciais fossem sistematicamente removidas ou falsificadas sem detecção;
- custos de adoção excluíssem justamente os grupos que se pretende proteger;
- múltiplas âncoras de confiança produzissem incompatibilidade insuperável;
- a perícia multimétodo não melhorasse confiabilidade ou contraditório.

Explicitar condições de falha impede que “ônus de realidade” se torne expressão retórica capaz de explicar qualquer resultado.

12.2 Limitações dogmáticas

O artigo concentra-se no processo civil. Regimes penal, eleitoral, trabalhista e administrativo possuem garantias, finalidades e regras próprias. A aplicação subsidiária do CPC ou o uso analógico das categorias requer análise específica. No processo penal, por exemplo, a dúvida e a distribuição do encargo estatal não podem ser reconfiguradas por uma matriz desenvolvida para litígios civis.

Quanto à relação entre o art. 373, § 1º, e a regra especial do art. 429, o artigo fixa apenas uma posição mínima: a dinamização não pode revogar silenciosamente a alocação legal, deve incidir sobre questão determinada e exige decisão autônoma que demonstre compatibilidade, necessidade e possibilidade de cumprimento. Não se pretende encerrar a controvérsia sobre todas as hipóteses em que a regra especial admitiria redistribuição. Tampouco se sustenta que o art. 436, isoladamente, discipline toda mídia digital; sua função é integrada aos arts. 422, 427–431 e 439–441.

A minuta da Rule 901(c) oferece comparação funcional, não argumento de autoridade. Diferenças entre o padrão norte-americano de autenticação, a divisão de funções entre juiz e

júri e o regime brasileiro impedem equivalência normativa. A convergência demonstra que o problema foi formulado independentemente; não prova a correção da solução nem autoriza importar o padrão “mais provável que não”.

12.3 Programa de validação

Uma agenda empírica brasileira poderia avançar em quatro etapas:

1. **estudo jurisprudencial** — mapear decisões que mencionam *deepfake*, IA generativa, autenticidade digital, art. 429 e art. 436, classificando objeto, fundamento, diligência, custo e resultado;
2. **vinhetas experimentais** — apresentar a magistrados, advogados, defensores e peritos cenários com diferentes níveis de especificidade e tipos de credencial, medindo decisões de perícia, distribuição do encargo e valoração;
3. **piloto institucional** — aplicar a matriz em conjunto controlado de processos ou procedimentos administrativos, com registro de tempo, custo, reversões e satisfação do contraditório;
4. **testes técnicos e distributivos** — avaliar ferramentas de proveniência e detecção em dispositivos, idiomas, grupos demográficos e condições de conectividade representativas do Brasil.

As métricas devem incluir não apenas acurácia técnica, mas falsos positivos, falsos negativos, custo, tempo, privacidade, acessibilidade, taxa de contestação, concordância entre avaliadores e distribuição dos encargos entre partes. Um protocolo que identifica mais *deepfakes*, mas inviabiliza prova genuína de grupos vulneráveis, não é institucionalmente superior.

13 CONCLUSÃO

O Radar Tecnológico nº 6 da ANPD estabelece base robusta para compreender geração, detecção, riscos, rastreabilidade e governança de *deepfakes*. Sua própria delimitação não exaustiva, porém, deixa espaço para desenvolver uma questão jurídico-probatória específica: o que acontece quando a possibilidade de falsificação se transforma em recurso para negar o verdadeiro?

A resposta exige tratar duas ameaças ao mesmo tempo. A primeira é a entrada do falso com aparência de autenticidade. A segunda é a retirada de eficácia do verdadeiro por uma dúvida de custo quase zero. Entre ambas surge o **ônus de realidade**: custo prático de

reconstruir origem, integridade e contexto que pode ser imposto a quem apresentou conteúdo genuíno antes mesmo de qualquer decisão formal sobre o ônus da prova.

O contraste norte-americano confirma a relevância do problema e restringe qualquer pretensão de novidade. A minuta da Rule 901(c) também adota um filtro inicial seguido de exigência reforçada ao proponente, mas permanece em estudo. O levantamento com 931 juízes encontrou apenas 15 experiências declaradas de contestação de *deepfake*, embora 82% dos respondentes sem experiência afirmassem que exigiriam demonstração inicial. Há convergência institucional sobre a utilidade de um limiar e, simultaneamente, insuficiência empírica para transformar a solução em dogma.

O direito brasileiro já contém um ponto de apoio decisivo. O art. 436, parágrafo único, não admite impugnação genérica de autenticidade ou falsidade. Essa regra deve conservar sentido diante da IA. A frase “pode ser *deepfake*” não individualiza o objeto, não formula hipótese concreta, não apresenta indício e não explica nexos. Por si só, não deve acionar perícia integral, perda de fé ou deslocamento de encargo.

A contestação legítima, por outro lado, não pode depender de laudo antecipado nem de domínio sobre o modelo generativo. A matriz proposta admite indícios técnicos, contextuais e de cadeia; calibra a exigência pela assimetria de acesso; separa autoria, origem, integridade, formação e contexto; e associa medidas progressivas à intensidade da dúvida.

O art. 429 permanece o ponto de partida para alocar o risco entre falsidade e impugnação de autenticidade. Preservação, exibição, cooperação e perícia dirigida podem ser determinadas sem alterar a regra de julgamento. Se a redistribuição do art. 373, § 1º, vier a ser necessária, deverá ocorrer por questão, mediante decisão fundamentada, sem surpresa e sem criar prova impossível. O ônus de realidade descreve um custo; não concede licença para inverter o ônus legal.

Deteção algorítmica e proveniência criptográfica ajudam, mas não certificam a verdade. Detectores erram, envelhecem e podem reproduzir vieses. Credenciais verificam vínculos e histórias declaradas; não demonstram que o evento ocorreu, que a declaração é verdadeira ou que o contexto é completo. Sua ausência não pode dividir o espaço probatório entre quem tem meios para certificar e quem precisa provar que merece ser ouvido.

A principal conclusão pode ser sintetizada:

A sociedade não preservará a verdade apenas aperfeiçoando a descoberta do falso. Precisarás impedir que a possibilidade do falso torne o verdadeiro processualmente inacessível.

O desafio não é criar uma autoridade infalível da realidade, mas condições públicas, auditáveis e inclusivas para especificar alegações, preservar arquivos, contestar métodos e justificar encargos. A evidência autêntica não merece imunidade nem prova indefinida de que o mundo ocorreu.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). *Deepfakes*. Brasília, DF: ANPD, 2026. 163 p. (Radar Tecnológico, n. 6). ISBN 978-65-82658-05-1. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/rt_deepfakes_anpd.pdf. Acesso em: 29 jul. 2026.
- BRASIL. *Lei nº 13.105, de 16 de março de 2015*. Código de Processo Civil. Brasília, DF: Presidência da República, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/13105compilada.htm. Acesso em: 29 jul. 2026.
- BRASIL. Superior Tribunal de Justiça (5. Turma). *Agravo Regimental no Habeas Corpus nº 828.054/RN*. Relator: Ministro Joel Ilan Paciornik, julgado em 23 abr. 2024, DJe 29 abr. 2024. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2024/02052024-Quinta-Turma-nao-aceita-como-provas-prints-de-celular-extraidos-sem-metodologia-adequada.aspx>. Acesso em: 29 jul. 2026.
- BRASIL. Superior Tribunal de Justiça (6. Turma). *Agravo Regimental no Habeas Corpus nº 1.014.212/ES*. Relator: Ministro Carlos Pires Brandão, julgado em 10 fev. 2026, DJEN 20 fev. 2026. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2026/09032026-Sexta-Turma-afasta-prisao-ate-conclusao-de-pericia-sobre-prints-de-WhatsApp-usados-como-prova.aspx>. Acesso em: 29 jul. 2026.
- CAPRA, Daniel J. Deepfakes reach the Advisory Committee on Evidence Rules. *Fordham Law Review*, New York, v. 92, n. 6, p. 2491–2506, 2024. Disponível em: <https://ir.lawnet.fordham.edu/flr/vol92/iss6/7/>. Acesso em: 29 jul. 2026.
- CHANDRA, Bilva; DUNIETZ, Jesse; ROBERTS, Kathleen; LEE, Yooyoung; FONTANA, Peter; AWAD, George. *Reducing risks posed by synthetic content: an overview of technical approaches to digital content transparency*. Gaithersburg: National Institute of Standards and Technology, 2024. (NIST AI 100-4). Atualizado em 8 abr. 2026. DOI: <https://doi.org/10.6028/NIST.AI.100-4>.
- CHESNEY, Robert; CITRON, Danielle Keats. Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*, Berkeley, v. 107, n. 6, p. 1753–1820, 2019. DOI: <https://doi.org/10.15779/Z38RV0D15J>.
- COALITION FOR CONTENT PROVENANCE AND AUTHENTICITY (C2PA). *C2PA and Content Credentials Explainer*. Version 2.2. [S. l.]: C2PA, 2025a. Disponível em: <https://spec.c2pa.org/specifications/specifications/2.2/explainer/Explainer.html>. Acesso em: 29 jul. 2026.
- COALITION FOR CONTENT PROVENANCE AND AUTHENTICITY (C2PA). *Content Credentials: C2PA Technical Specification*. Version 2.2, May 2025. [S. l.]: C2PA, 2025b. Disponível em: https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html. Acesso em: 29 jul. 2026.

- DIDIER JR., Fredie; BRAGA, Paula Sarno; OLIVEIRA, Rafael Alexandria de. *Curso de direito processual civil: teoria da prova, direito probatório, decisão, precedente, coisa julgada, processo estrutural e tutela provisória*. 21. ed. Salvador: JusPodivm, 2026. v. 2. 984 p. ISBN 978-85-442-7049-3.
- DORI, Caroline Lovison; CAMBI, Eduardo. Distribuição dinâmica do ônus da prova: exegese do art. 373, § 1º, CPC/2015. *Revista Eletrônica de Direito Processual*, Rio de Janeiro, v. 19, n. 2, p. 48–66, maio/ago. 2018. DOI: <https://doi.org/10.12957/redp.2018.33545>.
- FALLIS, Don. The epistemic threat of deepfakes. *Philosophy & Technology*, Dordrecht, v. 34, n. 4, p. 623–643, 2021. DOI: <https://doi.org/10.1007/s13347-020-00419-2>.
- FRICKER, Miranda. *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press, 2007. 188 p. DOI: <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>.
- GENOVA, Jandislei Antonio. *Autonomia decisória na era da IA: formação da vontade, influência cognitiva estrutural e responsabilidade jurídica em ambientes algorítmicos*. São Paulo: Editora Dialética, 2026. 156 p. ISBN 978-65-288-0467-2.
- HARRIS, Keith Raymond. Synthetic media detection, the wheel, and the burden of proof. *Philosophy & Technology*, Dordrecht, v. 37, art. 131, 2024. DOI: <https://doi.org/10.1007/s13347-024-00821-0>.
- MARTINEZ, Antonio Lopo. A prova sintética: deepfakes documentais e o futuro da verdade no processo brasileiro. *Revista Eletrônica de Direito Processual*, Rio de Janeiro, v. 27, n. 2, p. 65–90, maio/ago. 2026a. DOI: <https://doi.org/10.12957/redp.2026.97325>.
- MARTINEZ, Antonio Lopo. Jurisvision of deepfake financial documents: semiotic disruption and remediation in post-textual evidence. *International Journal for the Semiotics of Law*, Dordrecht, 3 jun. 2026b. DOI: <https://doi.org/10.1007/s11196-026-10514-w>.
- MATTHEWS, Taylor. Breaking the wheel, credibility, and hermeneutical injustice: a response to Harris. *Philosophy & Technology*, Dordrecht, v. 37, art. 138, 2024. DOI: <https://doi.org/10.1007/s13347-024-00828-7>.
- NEW YORK. Court of Appeals. *Matter of M.S. (M.H.)*. 2026 NY Slip Op 00825. Albany, 17 fev. 2026. Disponível em: https://www.nycourts.gov/reporter/3dseries/2026/2026_00825.htm. Acesso em: 29 jul. 2026.
- PLANTIN, Jean-Christophe; LAGOZE, Carl; EDWARDS, Paul N.; SANDVIG, Christian. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, London, v. 20, n. 1, p. 293–310, 2018. DOI: <https://doi.org/10.1177/1461444816661553>.
- RINI, Regina. Deepfakes and the epistemic backstop. *Philosophers' Imprint*, Ann Arbor, v. 20, n. 24, p. 1–16, 2020. Disponível em: <https://quod.lib.umich.edu/p/phimp/3521354.0020.024>. Acesso em: 29 jul. 2026.
- SCHIFF, Kaylyn Jackson; SCHIFF, Daniel S.; BUENO, Natália S. The liar's dividend: can politicians claim misinformation to evade accountability? *American Political Science Review*, Cambridge, v. 119, n. 1, p. 71–90, 2025. DOI: <https://doi.org/10.1017/S0003055423001454>.
- UNITED STATES. Judicial Conference of the United States. Advisory Committee on Evidence Rules. *Report to the Standing Committee*. Washington, DC, 17 May 2026. Disponível em: https://www.uscourts.gov/sites/default/files/document/advisory_committee_on_evidence_rules_may_2026.pdf. Acesso em: 29 jul. 2026.
- VACCARI, Cristian; CHADWICK, Andrew. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, London, v. 6, n. 1, p. 1–13, 2020. DOI: <https://doi.org/10.1177/2056305120903408>.