A DECISÃO NÃO COMEÇA NO CLIQUE

Ancoragem algorítmica, autonomia decisória e influência cognitiva estrutural

Jandislei Antonio Genova[1]

Resumo

A literatura sobre ancoragem demonstra que referências prévias podem deslocar julgamentos subsequentes. Em decisões assistidas por inteligência artificial, esse efeito pode ser ampliado pela seleção, personalização, repetição e mensuração contínua dos pontos de partida. O artigo investiga em que condições essa capacidade deixa de constituir influência episódica e passa a integrar uma arquitetura juridicamente relevante para a autonomia decisória. Adota-se pesquisa jurídico-dogmática, qualitativa, documental e interdisciplinar, apoiada em revisão narrativa auditável encerrada em 2 de agosto de 2026. O corpus reúne 44 fontes acadêmicas, doutrinárias, normativas e institucionais, incluindo estudos sobre ordem temporal do aconselhamento humano–IA e contraevidências relativas à experiência profissional. Não se reivindica novidade para a ancoragem algorítmica, para o juízo humano preliminar nem para intervenções de desancoramento. A contribuição consiste em reconstruir, no direito brasileiro, a *ancoragem cognitiva estrutural* como categoria analítica funcional e propor um protocolo heurístico não pontuado de oito eixos. O modelo separa mecanismo cognitivo, qualificação estrutural e consequência jurídica; não presume manipulação, vício do consentimento ou responsabilidade. A incidência jurídica depende de norma aplicável, materialidade, assimetria e prova do fato desencadeador. Propõem-se salvaguardas proporcionais para decisões híbridas, além de critérios de preservação, exibição e valoração da cronologia decisória. A matriz permanece sujeita a validação empírica, concordância entre avaliadores e análise de falsos positivos.

Palavras-chave: ancoragem algorítmica; inteligência artificial; influência cognitiva estrutural; autonomia decisória; supervisão humana.

1 Advogado e pesquisador independente. Especialista em Direito Civil e pós-graduado em Gestão em Saúde. Autor de Autonomia decisória na era da IA: formação da vontade, influência cognitiva estrutural e responsabilidade jurídica em ambientes algorítmicos (Editora Dialética, 2026).

Abstract

DECISION-MAKING DOES NOT BEGIN WITH THE CLICK: Algorithmic Anchoring, Autonomy in Decision-Making, and Structural Cognitive Influence

Research on anchoring shows that prior reference points may shift subsequent judgments. In AI-assisted decision-making, this effect may be amplified by the continuous selection, personalization, repetition, and measurement of starting points. This article examines when that capacity ceases to be episodic influence and becomes part of an architecture that is legally relevant to autonomy in decision-making. It adopts qualitative, documentary, interdisciplinary, and doctrinal legal research, supported by an auditable narrative review completed on 2 August 2026. The corpus comprises 44 academic, doctrinal, normative, and institutional sources, including studies on the timing of human–AI advice and counterevidence concerning professional expertise. No novelty is claimed for algorithmic anchoring, preliminary human judgment, or debiasing interventions. The contribution lies in reconstructing *structural cognitive anchoring* under Brazilian law as a functional analytical category and in proposing a non-scoring heuristic framework with eight dimensions. The framework separates the cognitive mechanism, structural qualification, and legal consequences; it does not presume manipulation, defective consent, or liability. Legal consequences depend on the applicable rule, materiality, asymmetry, and proof of the relevant triggering facts. The article proposes proportionate safeguards for hybrid decisions and criteria for preserving, disclosing, and evaluating decision-making chronology. The framework remains subject to empirical validation, inter-rater agreement, and false-positive analysis.

Keywords: algorithmic anchoring; artificial intelligence; structural cognitive influence; autonomy in decision-making; human oversight.

1 INTRODUÇÃO: A DECISÃO ANTES DA DECISÃO

Quando uma pessoa clica em “aceitar”, seleciona um candidato, confirma uma dose, fixa uma pena, escolhe um produto ou aprova uma operação de crédito, a decisão final costuma aparecer como o instante juridicamente visível da vontade. O clique, a assinatura ou a motivação escrita encerram a sequência e oferecem um autor identificável. Essa fotografia final pode ocultar a questão que a precede: quem determinou o ponto de partida a partir do qual a comparação foi realizada?

A ancoragem cognitiva descreve a influência exercida por uma informação inicial sobre estimativas e avaliações posteriores. O experimento clássico de Tversky e Kahneman

mostrou que números arbitrariamente produzidos deslocavam estimativas subsequentes (Tversky; Kahneman, 1974). A literatura posterior encontrou o efeito em avaliações econômicas, negociações, consumo, medicina e decisões judiciais, inclusive entre profissionais experientes (Furnham; Boo, 2011; Englich; Mussweiler; Strack, 2006). A âncora não precisa ordenar, proibir ou mentir. Basta reorganizar o campo comparativo e tornar cognitivamente mais custosa a distância em relação ao valor, rótulo ou recomendação inicialmente apresentado.

Sistemas de inteligência artificial não inventaram esse mecanismo. Eles modificam, porém, sua escala e sua governança. Uma plataforma pode decidir qual referência será exibida primeiro, selecionar o formato mais persuasivo, adaptar a recomendação a dados pessoais, repetir a exposição, medir a resposta e reposicionar a próxima referência. Em ambientes profissionais, o resultado pode aparecer como escore de risco, classificação, previsão, faixa recomendada, alerta ou resposta textual. Em interfaces de consumo, pode assumir a forma de preço anterior, opção "mais popular", ordem de resultados, estimativa de economia ou recomendação personalizada. Em todos esses casos, o valor do sistema não está apenas em prever a escolha: está também em participar da arquitetura que a forma.

Esse deslocamento leva à pergunta central:

Em que condições um ponto de referência produzido ou selecionado por uma arquitetura algorítmica deixa de constituir influência episódica e se converte em mecanismo juridicamente relevante de influência cognitiva estrutural?

A hipótese é que a origem algorítmica e a simples precedência de uma informação não bastam. A análise exige três planos sucessivos: (a) mecanismo cognitivo, demonstrado por deslocamento observado ou risco cientificamente plausível; (b) qualificação estrutural, identificada pelo controle persistente ou repetível da exposição e por déficit proporcional de independência; e (c) qualificação jurídica, dependente de norma aplicável, fato desencadeador, materialidade e padrão probatório. Personalização, repetição, opacidade, concentração de benefícios e transferência de riscos funcionam como agravantes, sem integrar obrigatoriamente todos os casos.

O argumento prolonga a investigação desenvolvida em *Autonomia decisória na era da IA*, na qual a influência cognitiva estrutural foi definida a partir da organização técnica e continuada das condições em que percepção, vontade e decisão se formam (Genova, 2026). O presente artigo isola um de seus mecanismos: o controle dos pontos de partida. A relação entre as categorias é de gênero e espécie funcional. Influência cognitiva estrutural descreve a

arquitetura mais ampla; ancoragem algorítmica descreve um efeito possível no julgamento; ancoragem cognitiva estrutural designa a configuração em que esse efeito é integrado a uma capacidade persistente de seleção, aprendizagem e intervenção.[2]

Três cautelas delimitam a tese. Primeiro, "algoritmo" não é agente moral autônomo: escolhas de projeto, objetivos, incentivos, dados e práticas organizacionais permanecem atribuíveis a pessoas e instituições.[3] Segundo, nem toda âncora produz erro. Pontos de referência informativos podem melhorar decisões quando possuem qualidade, pertinência e possibilidade de revisão. Terceiro, correlação entre exposição e escolha não basta para provar causalidade. O artigo distingue risco de ancoragem, deslocamento observado e exploração estrutural do efeito.

A contribuição é incremental e operacional. "Ancoragem algorítmica" já aparece em estudos sobre sistemas de recomendação, apoio automatizado e ordem temporal do aconselhamento. Também não são novos o juízo humano preliminar, o forçamento cognitivo ou a apresentação não simultânea da recomendação. A originalidade reivindicada é mais estreita: (a) separar mecanismo cognitivo, estrutura sociotécnica e consequência jurídica; (b) reconstruir essa relação no direito brasileiro; (c) propor um protocolo heurístico não pontuado de oito eixos e três núcleos cumulativos; e (d) vincular cada resposta jurídica a base normativa, fato desencadeador e padrão probatório próprios.

O texto está organizado em doze seções. Depois do método, reconstrói-se a literatura cognitiva e delimitam-se conceitos próximos. Em seguida, examina-se a transformação da âncora episódica em arquitetura adaptativa e confronta-se a proposta com evidências recentes de interação humano–IA. A análise jurídica articula Constituição, LGPD, CDC, Código Civil, CPC e disciplina do Conselho Nacional de Justiça, com contraste funcional do Regulamento Europeu de Inteligência Artificial. As seções finais apresentam a matriz, cenários, salvaguardas, consequências probatórias, limites e hipóteses de refutação.

2 MÉTODO, CORPUS E LIMITES

A pesquisa é jurídico-dogmática, qualitativa, documental e interdisciplinar. A reconstrução normativa parte de textos vigentes em 2 de agosto de 2026: Constituição da República, Lei Geral de Proteção de Dados Pessoais, Código de Defesa do Consumidor,

2 A expressão "ancoragem cognitiva estrutural" é empregada como categoria analítica funcional, sem reivindicação de prioridade lexical. A literatura já utiliza formulações próximas para descrever ancoragem produzida por sistemas, recomendações e ambientes algorítmicos. A contribuição reivindicada reside na delimitação jurídica e operacional proposta neste artigo.

3 Neste texto, expressões como "o sistema recomenda" descrevem função técnica ou organizacional e não atribuem consciência, vontade ou personalidade moral à tecnologia.

Código Civil, Código de Processo Civil, Resolução nº 615/2025 do Conselho Nacional de Justiça, com a alteração posterior pertinente, e Regulamento (UE) 2024/1689. A tomada de subsídios da ANPD é utilizada como contexto regulatório em formação, sem atribuição de força normativa. Fontes estrangeiras possuem função comparativa e não pressupõem identidade entre categorias, instituições ou efeitos jurídicos.[4]

O estado da arte foi construído por revisão narrativa orientada por problema, e não por revisão sistemática ou metanálise. Para torná-la auditável, registraram-se bases, expressões de busca, datas e critérios de seleção. Entre 1º e 2 de agosto de 2026, foram consultados ACM Digital Library, SpringerLink, ScienceDirect, Wiley Online Library, Cambridge Core, Portal de Periódicos IDP e catálogos editoriais, além dos repositórios oficiais Planalto, CNJ, ANPD e EUR-Lex. As buscas foram complementadas por rastreamento de referências e consulta por DOI.

Quadro 1 — Registro sintético do levantamento bibliográfico

Eixo	Expressões principais de busca	Critério de inclusão
Ancoragem e IA	"anchoring bias" AND ("AI-assisted decision-making" OR "algorithmic advice")	Estudos que isolassem referência inicial, conselho algorítmico ou deslocamento decisório.
Ordem temporal	("advice timing" OR "non-concurrent advice" OR "cognitive forcing") AND AI	Experimentos que comparassem aconselhamento anterior, simultâneo ou posterior ao juízo humano.
Supervisão humana	("human-in-the-loop" OR "human oversight") AND (anchoring OR overreliance OR "automation bias")	Pesquisas sobre dependência, erro, calibração e intervenção humana substancial.
Direito brasileiro	("decisão automatizada" OR "decisão híbrida") AND (LGPD OR supervisão humana); ancoragem AND inteligência artificial AND direito	Doutrina nacional, fontes normativas e documentos institucionais relacionados à autonomia, prova, consumo e proteção de dados.
Contraste comparado	AI Act AND "automation bias" AND "human oversight"	Fontes europeias destinadas apenas a contraste funcional.

Fonte: elaboração do autor. Buscas encerradas em 2 ago. 2026.

Foram incluídos trabalhos fundadores; revisões; estudos originais com metadados verificáveis; pesquisas com profissionais ou decisões de consequência elevada; estudos sobre aconselhamento, ordem temporal e interação humano–IA; doutrina jurídica brasileira pertinente; e fontes normativas ou institucionais oficiais. Excluíram-se duplicatas, textos sem autoria ou metadados verificáveis, peças meramente promocionais e materiais sem relação direta com a pergunta. O corpus final contém 44 entradas: 28 trabalhos acadêmicos estrangeiros, sete obras doutrinárias brasileiras independentes, oito fontes normativas ou

4 O Regulamento Europeu de IA é utilizado como contraste funcional sobre supervisão humana e viés de automação, não como fonte diretamente aplicável ao regime brasileiro.

institucionais e um livro anterior do autor. A classificação organiza o corpus, mas não transforma a revisão narrativa em levantamento exaustivo.

O método impõe limites. Não se realizou experimento próprio, levantamento sistemático de jurisprudência brasileira, auditoria de interface ou pesquisa de campo com consumidores, magistrados, médicos, recrutadores ou gestores. Estudos estrangeiros não demonstram, por si, como grupos brasileiros reagirão em contextos reais. Resultados controlados variam com experiência, risco, incentivo, pressão temporal, qualidade da recomendação e cultura organizacional. O artigo oferece uma categoria analítica e um protocolo heurístico de investigação; não afirma prevalência empírica nacional nem validação do instrumento.

Também se evita transformar toda influência em manipulação. A vida social depende de referências, conselhos, padrões e comparações. Uma estimativa tecnicamente sólida pode reduzir erro e distribuir conhecimento. O problema jurídico é relacional e contextual: quem define a referência, com qual fundamento, sob quais assimetrias, com que consequências e mediante quais possibilidades de compreender, divergir e recorrer.

Para preservar essa distinção, o artigo utiliza três níveis de afirmação:

1. risco arquitetural: a interface e o processo possuem características capazes de favorecer ancoragem;

2. efeito observado: existe evidência experimental, quase experimental, comportamental ou documental de deslocamento em relação a linha de base adequada;

3. qualificação estrutural: o efeito ou risco substancial integra capacidade persistente ou repetível de controlar exposições sob déficit proporcional de independência.

A consequência jurídica constitui etapa posterior: depende de norma aplicável, fato desencadeador, materialidade, nexo e padrão probatório. Essa separação impede presumir manipulação sempre que uma IA apresenta recomendação ou considerar livre toda decisão apenas porque uma pessoa formalmente pôde clicar em opção diferente.

3 ANCORAGEM COGNITIVA: MECANISMOS, ROBUSTEZ E FRONTEIRAS

3.1 O ponto de referência e o deslocamento subsequente

Na formulação original, a ancoragem integra o conjunto de heurísticas utilizado em julgamentos sob incerteza. Pessoas expostas a um número inicial tendem a produzir estimativas posteriores mais próximas dele do que pessoas expostas a referência diversa (Tversky; Kahneman, 1974). A robustez do fenômeno foi documentada em tarefas e domínios

heterogêneos, embora sua magnitude dependa do desenho experimental, da plausibilidade da referência, do conhecimento e do modo de resposta (Furnham; Boo, 2011).

Não existe explicação única para todos os casos. O modelo de ajuste insuficiente sustenta que o decisor parte do valor inicial e se afasta dele menos do que seria necessário. A teoria da acessibilidade seletiva enfatiza que a comparação com a âncora ativa informações compatíveis com ela, tornando-as mais disponíveis para o julgamento posterior. Há ainda explicações pragmáticas, escalares e perceptivas. A evidência de ancoragem perceptiva mostra que referências podem alterar não apenas cálculos deliberados, mas a própria representação comparativa de estímulos (Jain; Nayakankuppam; Gaeth, 2021).

Essa diversidade importa juridicamente. Se o efeito decorrer apenas de ajuste consciente, advertências e tempo adicional talvez sejam suficientes. Se a referência reorganizar a acessibilidade de informações, uma explicação posterior pode chegar tarde: o campo de análise já foi estreitado. E, quando a âncora é apresentada como conselho competente, sua influência pode combinar mecanismos de ancoragem, confiança e deferência.

O estudo de Englich, Mussweiler e Strack (2006) é particularmente relevante para o direito. Profissionais do sistema de justiça foram influenciados por demandas sancionatórias irrelevantes, inclusive quando a referência derivava de mecanismo aleatório. O resultado não autoriza generalizar que toda decisão judicial é ancorada.

Contraevidência posterior impede conclusão uniforme. Em experimento pré-registrado com leigos, estudantes e profissionais do direito, Teichman, Zamir e Ritov (2023) encontraram efeito de âncora entre leigos e estudantes, mas não entre os profissionais nas condições pesquisadas. O resultado não demonstra imunidade profissional; mostra que experiência, significado da referência e contexto podem moderar o efeito. A hipótese deste artigo, portanto, não depende da premissa de que toda recomendação algorítmica ancorará qualquer especialista.

3.2 Âncora externa, âncora autogerada e conselho

Uma referência pode ser externa — apresentada por outra pessoa, documento, interface ou sistema — ou autogerada pelo próprio decisor. Essa diferença altera os mecanismos e as estratégias de correção. Advertir sobre o viés pode funcionar de modo desigual, e incentivos para precisão não garantem neutralização. Em ambientes digitais, a distinção é ainda mais difícil porque preferências passadas do próprio usuário podem retornar

como recomendação externa. O sistema transforma histórico comportamental em ponto de partida e o reapresenta com aparência de previsão independente.

Conselho e âncora também não são sinônimos. Um conselho possui pretensão informativa: supõe-se que o conselheiro detenha conhecimento relevante. Uma âncora arbitrária pode influenciar sem essa pretensão. Hütter e Fiedler (2019) demonstram que o processamento de conselho genuíno e de âncoras arbitrárias não é idêntico. Em decisões assistidas por IA, uma mesma saída pode exercer as duas funções: oferecer informação e, pela precedência, fixar uma faixa comparativa. A avaliação jurídica deve examinar sua qualidade sem ignorar o efeito posicional.

3.3 Conceitos próximos que não devem ser confundidos

Quadro 2 — Delimitação entre mecanismos próximos

Conceito	Operação predominante	Relação com a ancoragem	Risco de confusão
Ancoragem	Um ponto de referência desloca avaliação subsequente.	É o mecanismo central examinado.	Chamar de âncora toda informação apresentada primeiro.
Primazia	Elementos iniciais recebem peso desproporcional na memória ou impressão.	Pode favorecer a força da âncora, mas não exige ajuste numérico.	Tratar ordem temporal como prova suficiente de ancoragem.
Enquadramento	A formulação do problema altera sua interpretação.	Pode definir o significado da referência e das alternativas.	Confundir mudança de descrição com deslocamento em torno de um ponto.
Default	Uma opção é previamente selecionada e exige ação para ser alterada.	Pode atuar como âncora normativa ou prática.	Pressupor que todo *default* opera pelo mesmo mecanismo cognitivo.
Viés de automação	Há uso excessivo, omissão de verificação ou deferência à saída automatizada.	A saída pode ancorar a decisão e contribuir para a deferência.	Reduzir toda confiança em automação a ancoragem.
Aconselhamento	Informação de terceiro é incorporada ao julgamento.	O conselho pode ser informativo e simultaneamente funcionar como âncora.	Considerar influência como irracional mesmo quando melhora a decisão.
Manipulação	Influência oculta ou indevida subverte a capacidade decisória.	Ancoragem pode ser explorada manipulativamente.	Presumir intenção oculta apenas a partir do efeito cognitivo.

Fonte: elaboração do autor com base em Tversky e Kahneman (1974), Hütter e Fiedler (2019), Parasuraman e Manzey (2010) e Susser, Roessler e Nissenbaum (2019).

A distinção mais importante envolve viés de automação. Ele costuma aparecer quando usuários aceitam recomendações automatizadas ou deixam de buscar informação incompatível (Mosier; Skitka, 1999; Parasuraman; Manzey, 2010). A ancoragem pode ser um dos mecanismos dessa deferência, mas não o único. Confiança institucional, autoridade percebida, carga de trabalho, aversão à responsabilidade, desenho da interface e cultura organizacional também influenciam. A categoria jurídica precisa resistir à tentação de explicar fenômenos heterogêneos com um único rótulo.

3.4 Anterioridade direta e limite da originalidade

A literatura sobre decisões assistidas por IA já desenvolveu intervenções próximas às salvaguardas temporais discutidas neste artigo. Buçinca, Malaya e Gajos (2021) testaram funções de forçamento cognitivo destinadas a reduzir dependência excessiva. Rastogi *et al.* (2022) modelaram vieses em decisões assistidas, com atenção à ancoragem e a estratégias temporais de desancoramento. Agudo *et al.* (2024) manipularam o momento em que participantes recebiam apoio algorítmico em julgamentos simulados e observaram maior prejuízo quando recomendação incorreta antecedia o juízo próprio. Kokje *et al.* (2026), em três experimentos pré-registrados, encontraram que apresentações não simultâneas podem reduzir dependência excessiva, mas também produzir subutilização de recomendações corretas, sem ganho uniforme de desempenho.

Há convergência igualmente direta no plano judicial. Strikaitė-Latušinskaja (2026) examina recomendações algorítmicas como pontos de referência dominantes capazes de estreitar discricionariedade e individualização. Esses trabalhos impedem reivindicar novidade para o diagnóstico geral, para a precedência temporal ou para o juízo humano preliminar.

A diferença específica deste artigo reside em outro plano: integrar a evidência comportamental a uma reconstrução do direito brasileiro; separar mecanismo, estrutura e consequência jurídica; e organizar fatos, registros e respostas num protocolo probatório não pontuado. A anterioridade funciona, assim, como limite e teste de diferenciação, não como negação da contribuição.

4 DA ÂNCORA EPISÓDICA À ARQUITETURA ADAPTATIVA

4.1 O que a IA efetivamente acrescenta

Uma âncora tradicional pode ser casual, estática e igual para todos. Sistemas digitais permitem uma operação diferente. Eles podem estimar quais referências possuem maior probabilidade de alterar a ação de um usuário, testar variantes, escolher o momento da apresentação e repetir a intervenção. O ciclo combina observação, previsão e modificação: o ambiente coleta sinais da conduta; o modelo estima respostas; a interface seleciona uma referência; a reação retorna como novo dado.

Yeung (2017) denominou *hypernudge* a forma de regulação por design baseada em dados, dinamismo e atualização contínua da arquitetura da escolha. Calo (2014) mostrou que mercados digitais ampliam a capacidade empresarial de identificar e explorar vulnerabilidades em nível individual. Susser, Roessler e Nissenbaum (2019) situam a manipulação on-line na

influência oculta que subverte o poder decisório. Essas construções são mais amplas do que a ancoragem, mas explicam o ambiente em que uma referência deixa de ser evento isolado.

A passagem do episódico ao estrutural não depende de um "algoritmo mal-intencionado". Um sistema otimizado para conversão, produtividade, redução de perdas ou uniformidade decisória pode aprender que determinadas faixas, rótulos e recomendações aumentam a métrica, ainda que ninguém tenha programado explicitamente a exploração de um viés. Schmauder *et al.* (2023) demonstram por que intervenções algorítmicas demandam supervisão interdisciplinar: uma estratégia pode ser eficaz para o objetivo operacional e, simultaneamente, utilizar processos cognitivos não identificados pelos próprios desenvolvedores.

4.2 Proposta conceitual

Propõe-se a seguinte definição funcional:

Ancoragem cognitiva estrutural *é a configuração sociotécnica na qual o agente que organiza o ambiente decisório controla, de modo persistente ou repetível, a seleção, a forma ou a temporalidade de referências iniciais capazes de deslocar materialmente avaliações subsequentes, sob condições de independência proporcionalmente insuficientes.*

"Capaz de deslocar" não significa determinação inevitável. O conceito é probabilístico. Pessoas podem rejeitar a referência, sistemas podem melhorar decisões e efeitos podem ser nulos. A qualificação "estrutural" não decorre do tamanho da empresa nem do uso de aprendizado de máquina, mas da combinação entre controle relevante, repetibilidade e déficit proporcional de independência.

Para evitar circularidade, a análise é feita em etapas. Primeiro, investiga-se o mecanismo cognitivo: exposição, ordem e deslocamento ou risco plausível. Depois, examina-se a estrutura: controle da arquitetura, repetibilidade, assimetria e contestabilidade. Somente então ocorre a qualificação jurídica, mediante norma específica, consequência material, nexo e padrão probatório. Benefício econômico e transferência de risco agravam a inferência, mas não integram necessariamente a definição. Neste artigo, "agente organizador da arquitetura" designa quem controla seleção, interface ou fluxo; "controlador" fica reservado ao sentido técnico da LGPD quando houver tratamento de dados pessoais.[5]

5 Para evitar ambiguidade com o art. 5º, VI, da LGPD, "controlador" é utilizado apenas quando se trata do agente de tratamento que decide sobre dados pessoais. Nos demais contextos, emprega-se "agente organizador da arquitetura".

Quadro 3 — Tipologia da ancoragem em ambientes algorítmicos

Nível	Configuração	Evidência mínima	Consequência analítica
Referência informativa	Valor ou recomendação pertinente, acompanhado de fundamento e alternativas.	Validação técnica e pertinência ao caso.	Pode melhorar a decisão; não se presume viés.
Ancoragem episódica	Referência inicial influencia um julgamento específico.	Comparação entre decisões expostas e linha de base adequada.	Efeito cognitivo localizado.
Ancoragem algorítmica	Saída produzida ou selecionada por sistema funciona como ponto de partida.	Cronologia da exposição e deslocamento compatível com a referência.	Exige examinar qualidade, interface e supervisão.
Ancoragem cognitiva estrutural	O agente organizador controla, de modo persistente ou repetível, referência material sob independência insuficiente	. Convergência entre mecanismo cognitivo e condições estruturais.	Justifica investigação jurídica posterior, sem efeito normativo automático.

Fonte: elaboração do autor.

A tipologia protege a análise contra dois erros. O primeiro é moralizar toda influência: recomendações competentes são indispensáveis em medicina, gestão e justiça. O segundo é aceitar a assinatura humana como prova de independência: uma decisão pode ser formalmente humana e materialmente pré-estruturada por um sistema.

5 EVIDÊNCIA EMPÍRICA EM INTERAÇÕES HUMANO–IA

5.1 Recomendações, desempenho e persistência

Estudos recentes fornecem evidência de que saídas algorítmicas podem operar como âncoras, mas os resultados não são uniformes. Carter e Liu (2025) realizaram dois experimentos controlados com 775 gestores e encontraram efeitos de referências altas e baixas em avaliações de desempenho assistidas por IA. A estratégia de “considerar o oposto” reduziu a influência nas condições estudadas. Köcher *et al.* (2019) identificaram ancoragem em atributos de produtos recomendados. Narayanan, Somasundaram e Seifert (2025) observaram influência persistente de apoio algorítmico avesso ao risco em decisões de estoque, inclusive depois da retirada do conselho.

Em seleção profissional, Cecil *et al.* (2024) conduziram cinco experimentos pré-registrados com 1.403 participantes. Recomendações corretas melhoraram decisões, mas recomendações incorretas foram frequentemente seguidas; diferentes explicações não reduziram de modo consistente a dependência indevida. A conclusão normativa não é que explicabilidade seja inútil, mas que uma razão visual ou textual não garante supervisão independente.

5.2 Ordem temporal, forçamento cognitivo e contraefeitos

Buçinca, Malaya e Gajos (2021), em experimento com 199 participantes, compararam funções de forçamento cognitivo com formas usuais de aconselhamento explicável. Intervenções que exigiam engajamento analítico reduziram a dependência excessiva, mas aumentaram esforço e foram menos preferidas. Rastogi *et al.* (2022) trataram ancoragem, confirmação e disponibilidade como componentes distintos da decisão assistida e testaram estratégias temporais de desancoramento.

Agudo *et al.* (2024) realizaram dois experimentos simulando julgamentos sobre responsabilidade criminal. Recomendações incorretas apresentadas antes do juízo próprio reduziram a precisão; formular avaliação inicial ajudou, mas não eliminou a influência posterior. Em três experimentos pré-registrados de reconhecimento facial, Kokje *et al.* (2026) encontraram que aconselhamento não simultâneo pode reduzir a aceitação de erros, porém também diminuir a utilização de recomendações corretas. Não houve melhora uniforme do desempenho conjunto.

Esses resultados sustentam duas conclusões simultâneas. A ordem da informação é variável causal relevante; e nenhuma sequência constitui salvaguarda universal. Juízo preliminar, apresentação sob demanda ou aconselhamento condicional precisam ser calibrados pelo domínio, pela qualidade do sistema e pelo risco de subutilização.

5.3 Saúde e formas de apresentação

Em decisões médicas, Adam *et al.* (2022) estudaram 438 clínicos e 516 não especialistas em tarefas de emergência. Recomendações prescritivas enviesadas deslocaram decisões; apresentação descritiva de sinais relevantes mitigou parte do efeito. Em vez de uma conclusão única antes da avaliação, o sistema pode fornecer elementos que preservem a formação inicial do juízo humano.

A mesma informação pode produzir efeitos diferentes conforme apareça como primeiro veredito, segunda opinião, faixa de incerteza ou conjunto de evidências. Supervisão humana não é apenas presença de uma pessoa ao final do fluxo; compreende a sequência pela qual pessoa e sistema formam e revisam julgamentos.

5.4 Laços de retroalimentação e aprendizagem do viés

Glickman e Sharot (2025) demonstraram, em série de experimentos com 1.401 participantes, que interações repetidas com sistemas enviesados podem amplificar julgamentos humanos sem percepção equivalente da influência. Ikeda (2024), com 1.925 participantes, encontrou influência de conselhos do ChatGPT em diferentes domínios.

Nguyen (2024) mostrou que modelos de linguagem também podem ser suscetíveis a referências em tarefas financeiras.

É essencial separar dois objetos: um modelo pode ser ancorado pelo *prompt*; um humano pode ser ancorado pela resposta do modelo. O primeiro envolve comportamento do sistema; o segundo, formação da decisão humana. Eles podem compor ciclo de retroalimentação, mas exigem evidências diferentes.

Quadro 4 — Evidência selecionada sobre referências e apoio algorítmico

Estudo	Contexto	Resultado pertinente	Limite para este artigo
Buçinca, Malaya e Gajos (2021)	Decisão assistida e explicações	Forçamento cognitivo reduziu dependência excessiva, com custos de esforço e preferência.	Uma tarefa controlada não valida salvaguarda universal.
Rastogi *et al.* (2022)	Decisão assistida por IA	Modelou vieses e testou intervenção temporal de desancoramento.	Resultados dependem da tarefa e do modelo comportamental.
Agudo *et al.* (2024)	Julgamentos simulados	Apoio incorreto anterior ao juízo próprio reduziu precisão.	Participantes leigos e sistema simulado limitam validade externa.
Kokje *et al.* (2026)	Reconhecimento facial	Atraso pode reduzir sobreutilização e aumentar subutilização, sem ganho uniforme.	Não há solução temporal única para todos os domínios.
Carter e Liu (2025)	Avaliação de desempenho	Âncoras deslocaram avaliações; “considerar o oposto” mitigou efeitos.	Experimentos não reproduzem toda a organização.
Cecil *et al.* (2024)	Seleção profissional	Conselho incorreto prejudicou decisões; explicações não resolveram dependência.	Tarefa simulada e amostra limitam generalização.
Adam *et al.* (2022)	Emergência médica	Recomendação prescritiva enviesada influenciou; forma descritiva mitigou parte.	Cenário experimental não substitui prática longitudinal.
Teichman, Zamir e Ritov (2023)	Decisão jurídica	Âncora irrelevante não produziu efeito significativo entre profissionais nas condições testadas.	Não demonstra imunidade a âncoras relevantes ou algorítmicas.
Glickman e Sharot (2025)	Interações repetidas	Laços humano–IA amplificaram pequenos vieses.	Não isola ancoragem como mecanismo único.

Fonte: elaboração do autor a partir dos estudos indicados.

5.5 O que a evidência permite e não permite afirmar

O conjunto sustenta proposições moderadas: saídas de IA podem deslocar decisões; ordem e forma de apresentação importam; explicações isoladas podem falhar; repetição pode produzir persistência; e intervenções contra sobreutilização podem gerar subutilização. Não permite concluir que toda recomendação é manipulativa, que especialistas sempre obedecem ou que IA necessariamente piora resultados.

Também não se deve confundir diferença em relação à linha de base com dano. Uma recomendação correta pode deslocar o juízo para melhor. Estudos heterogêneos não validam,

em conjunto, a categoria jurídica proposta nem seus oito eixos. A matriz adiante organiza investigação; sua utilidade e confiabilidade dependem de testes próprios, inclusive contra casos negativos e contraevidências.

6 RECONSTRUÇÃO JURÍDICA NO DIREITO BRASILEIRO

6.1 Autonomia, dignidade e formação da vontade

A Constituição protege dignidade, liberdade, intimidade e o direito fundamental à proteção de dados pessoais, especialmente nos arts. 1º, III, e 5º, II, X e LXXIX. Esses fundamentos não consagram vontade imune a influências, condição inexistente na vida social. Protegem a pessoa contra estruturas de poder que esvaziem sua capacidade prática de compreender, escolher e contestar.

No plano civil, a ancoragem não constitui novo vício autônomo. Sua eventual relevância deve ser reconduzida aos requisitos do negócio, à interpretação conforme boa-fé, aos defeitos da vontade, ao abuso e à responsabilidade previstos, entre outros, nos arts. 104, 113, 138 a 165, 187 e 421 a 422 do Código Civil. A reconstrução civil contemporânea exige individualizar o instituto aplicável, em vez de converter achado psicológico em causa geral de invalidade (Schreiber, 2026).

A autonomia decisória é empregada como categoria reconstrutiva: descreve condições materiais de formação e revisão da vontade, sem pretender substituir categorias legais. A consequência depende de pertinência, materialidade, nexo e enquadramento normativo demonstrados no caso.

6.2 LGPD e a zona das decisões híbridas

A LGPD oferece fundamentos específicos nos arts. 6º, 8º, 9º, 18, 20 e 38. Finalidade, adequação, necessidade, transparência, prevenção, não discriminação e responsabilização incidem sobre tratamentos utilizados para perfilamento, recomendação e suporte decisório. Consentimento, quando for a base legal pertinente, não deve ser tratado como autorização autossuficiente em ambiente marcado por assimetrias de poder e informação (Bioni, 2021; Doneda, 2021).

O art. 20 disciplina decisões tomadas unicamente com base em tratamento automatizado que afetem interesses do titular. O problema da ancoragem surge também em decisões híbridas: o sistema recomenda e uma pessoa formalmente decide. A redação legal não autoriza aplicar automaticamente o regime do art. 20 a todo apoio algorítmico, mas

tampouco torna irrelevantes os princípios, direitos de acesso, correção e informação ou a responsabilização do agente de tratamento.[6]

A tomada de subsídios realizada pela ANPD entre 2024 e 2025 incluiu expressamente a revisão de decisões automatizadas, os limites do art. 20, direitos dos titulares e boas práticas. Seus resultados constituem contexto regulatório em formação, e não regulamento vinculante (Agência Nacional de Proteção de Dados, 2025). A literatura brasileira sobre discriminação algorítmica igualmente demonstra que perfilamento e decisão devem ser avaliados à luz de fundamentos jurídicos e efeitos concretos, não apenas pela aparência de neutralidade técnica (Mendes; Mattiuzzo, 2019).

Em fluxo híbrido, a questão probatória é substancial: o humano formou juízo próprio, teve tempo e competência, acessou informações divergentes, pôde rejeitar a saída sem sanção e registrou razões independentes? A assinatura do analista não responde, sozinha, a essas perguntas.

6.3 Relações de consumo e vulnerabilidade

Os arts. 4º, I, 6º, II e III, 39, IV e V, 46 e 51 do CDC protegem vulnerabilidade, liberdade de escolha, informação adequada e equilíbrio contratual. A dogmática consumerista não proíbe persuasão, recomendações ou preços de referência; controla práticas que explorem vulnerabilidade, ocultem elementos relevantes ou convertam assimetria informacional em vantagem incompatível com boa-fé e equilíbrio (Marques, 2025; Miragem, 2024).

Preços anteriores, descontos, ordenação e recomendações podem ser legítimos. Tornam-se juridicamente relevantes quando a referência é falsa, incomparável, materialmente incompleta, adaptada para explorar vulnerabilidade ou apresentada de modo a tornar alternativas relevantes impraticáveis. A coleta de dados acrescenta assimetria: o fornecedor pode estimar limiares de preço, urgência ou confiança, enquanto o consumidor desconhece que a referência foi selecionada em função do perfil.

Em litígios, a capacidade adaptativa pode justificar pedido delimitado de exibição de versões de interface, critérios de segmentação, histórico de testes e registros de exposição, observadas proporcionalidade, segredo empresarial e proteção de terceiros. Não se presume abusividade apenas porque houve personalização.

6 “Decisão híbrida” designa o fluxo no qual uma saída automatizada informa, recomenda, prioriza ou estrutura uma decisão formalmente atribuída a pessoa natural. A intensidade da participação humana deve ser verificada no caso concreto.

6.4 Boa-fé, dever de informar e responsabilidade

A organização intencional do ambiente para obter adesão previsível pode, conforme o caso, caracterizar violação da boa-fé, abuso ou inadimplemento do dever de informar. A conclusão não decorre da ancoragem isolada: exige demonstrar relação jurídica, dever aplicável, comportamento contraditório ou desleal, materialidade e efeito juridicamente relevante (Schreiber, 2026).

O dever de informar não se satisfaz necessariamente pela acumulação de texto. Informação adequada deve ser tempestiva, inteligível e relacionada ao risco. Explicação exibida apenas depois da referência ou escondida em fluxo secundário pode existir formalmente e falhar funcionalmente, especialmente em relações de consumo.

Responsabilidade civil exige dano, nexo e fundamento de imputação conforme o regime aplicável. A ancoragem pode integrar a prova do nexo ou da falha do serviço, mas raramente os demonstra sozinha. Cadeias decisórias possuem múltiplas causas; por isso, o protocolo proposto serve para formular questões e preservar evidências, não para substituir os requisitos jurídicos.

6.5 Processo, prova e cronologia decisória

Os arts. 369, 373, § 1º, e 396 a 400 do CPC oferecem bases distintas para produção da prova, distribuição dinâmica do ônus e exibição. Essas técnicas não se confundem. A redistribuição depende de decisão fundamentada, dificuldade concreta e oportunidade de cumprimento; a exibição exige delimitação do documento ou categoria; a valoração permanece vinculada ao conjunto probatório (Didier Jr.; Braga; Oliveira, 2026).

Em controvérsia sobre ancoragem, o objeto probatório deve ser decomposto:

1. qual referência foi apresentada;

2. em que momento, versão e formato;

3. quais dados determinaram sua seleção;

4. quais alternativas estavam disponíveis e visíveis;

5. se havia linha de base ou juízo preliminar independente;

6. quais taxas de aceitação, rejeição e alteração foram registradas;

7. como divergências eram tratadas pela organização;

8. qual consequência material ocorreu e quem controlava a evidência.

O agente organizador normalmente possui logs, versões, testes de interface e métricas; o afetado conhece sua experiência, mas não a infraestrutura. Essa assimetria pode sustentar

preservação, exibição e, em situação própria, distribuição dinâmica, sem que qualquer delas antecipe responsabilidade.

6.6 Uso de IA no Judiciário e supervisão humana

A Resolução nº 615/2025 do CNJ estabelece diretrizes para desenvolvimento, governança e uso de IA no Poder Judiciário. Entre seus eixos estão participação e supervisão humanas, contestabilidade, rastreabilidade, explicabilidade, capacitação e controle de riscos. A disciplina veda sistemas que eliminem revisão humana ou produzam dependência absoluta.

A ancoragem oferece conteúdo operacional a esses deveres. Magistrado ou servidor pode conservar competência formal e ser condicionado por recomendação exibida antes da leitura dos autos, por *ranking*, minuta ou faixa de sanção. Strikaitė-Latušinskaja (2026) demonstra, em perspectiva comparada, como recomendações podem tornar-se referências dominantes e estreitar discricionariedade. Supervisão efetiva requer ordem proporcional de apresentação, acesso a elementos incompatíveis, registro de alteração e ausência de incentivos que punam divergência.

Não se defende impedir apoio algorítmico. Sistemas podem localizar precedentes e aumentar consistência. O risco surge quando eficiência é medida por adesão ou velocidade de confirmação, convertendo a presença humana em carimbo. Koulu (2020) adverte que supervisão pode tornar-se invólucro procedimental quando se preserva o gesto humano e se removem as condições materiais de discricionariedade.

6.7 Contraste funcional com o Regulamento Europeu de IA

O art. 14 do Regulamento (UE) 2024/1689 exige supervisão efetiva em sistemas de alto risco. O supervisor deve compreender capacidades e limitações, reconhecer tendência à confiança automática ou excessiva, interpretar resultados e poder ignorar, reverter ou interromper o sistema. A norma menciona expressamente o viés de automação.

Laux e Ruschemeier (2025) observam que exigir consciência do viés é insuficiente se design e contexto organizacional continuarem produzindo deferência. Saber que ancoragem existe não desfaz necessariamente sua influência. Salvaguardas precisam alcançar sequência, forma da recomendação, incentivos e registros de divergência.

O contraste não autoriza transplante automático ao Brasil. Demonstra apenas convergência funcional: supervisão significativa precisa ser projetada contra dependência excessiva, e não presumida pela presença de uma pessoa.

7 PROTOCOLO HEURÍSTICO DE ANCORAGEM COGNITIVA ESTRUTURAL

O instrumento abaixo organiza perguntas, evidências e respostas proporcionais. É protocolo heurístico não pontuado: não constitui escala psicométrica, cálculo certificado, teste causal ou presunção de ilicitude. Seus eixos devem ser avaliados conjuntamente e os resultados precisam declarar incertezas, hipóteses rivais e evidências ausentes.

Quadro 5 — Protocolo heurístico de oito eixos

Eixo	Pergunta diagnóstica	Evidência relevante	Sinal de agravamento
Controle da exposição	Quem escolhe referência, formato e momento?	Regras de negócio, documentação e versões da interface.	O beneficiário controla integralmente o ponto de partida.
Precedência temporal	O decisor forma avaliação antes ou depois da saída?	Fluxo de telas, marcações de tempo e protocolo.	A conclusão aparece antes das evidências.
Personalização	A referência varia conforme perfil ou vulnerabilidade inferida?	Variáveis, segmentação e critérios de recomendação.	Ajuste individual orientado à conversão ou conformidade.
Repetição e persistência	A exposição é reiterada ou sobrevive à retirada?	Histórico, séries temporais e testes de remoção.	Ciclo que aprende com aceitação e resistência.
Opacidade e alternativas	Origem, limites e opções excluídas são compreensíveis?	Avisos, documentação, testes e alternativas.	Explicação formal sem inteligibilidade ou opção real.
Materialidade	A referência afeta direito, renda, saúde, liberdade ou oportunidade?	Resultado, magnitude, reversibilidade e grupos afetados.	Consequência grave ou difícil de reparar.
Benefício e risco	Quem obtém benefício e quem suporta o erro?	Métricas, remuneração e alocação de perdas.	Benefício concentrado e risco transferido.
Contestabilidade	É possível rejeitar, obter segunda opinião e recorrer sem retaliação?	Divergências, procedimentos, tempos e revisões.	Rejeição técnica possível, mas institucionalmente punida.

Fonte: elaboração do autor.

7.1 Núcleos obrigatórios, agravantes e qualificação jurídica

Para evitar deriva conceitual, três núcleos são cumulativos: domínio relevante do ponto de partida; capacidade de deslocamento material, observada ou cientificamente plausível; e déficit proporcional de independência. Personalização, repetição, opacidade, benefício e transferência de risco são agravantes. A consequência jurídica somente aparece numa etapa autônoma.

Quadro 6 — Correspondência entre diagnóstico, estrutura e direito

Camada	Elementos	Função	Consequência da ausência
Mecanismo cognitivo	Exposição, precedência e deslocamento observado ou risco plausível.	Identificar ancoragem ou risco de ancoragem.	Não se afirma efeito; permanece hipótese de design.
Núcleo estrutural 1	Controle relevante da seleção, forma ou temporalidade.	Atribuir domínio do ponto de partida.	Permanece influência episódica ou externa.
Núcleo estrutural 2	Consequência material e capacidade de deslocamento.	Excluir referências triviais.	Não há qualificação estrutural para o caso individual.

Camada	Elementos	Função	Consequência da ausência
Núcleo estrutural 3	Contestação, alternativas, tempo, competência ou incentivos insuficientes.	Verificar independência substancial.	Há conselho ou apoio contestável, não captura estrutural.
Agravantes	Personalização, repetição, opacidade, benefício concentrado e risco transferido.	Aumentar força da inferência e exigência de governança.	A categoria ainda pode existir, com menor intensidade.
Qualificação jurídica	Norma, fato desencadeador, nexo, prova e consequência.	Definir resposta jurídica possível.	O diagnóstico não produz efeito normativo.

Fonte: elaboração do autor.

7.2 Diferença entre risco, efeito e exploração

O desenho pode ser de alto risco antes de dano observado. Nesse plano preventivo, plausibilidade cognitiva, controle e gravidade potencial podem justificar testes e registros. Para afirmar efeito, exige-se comparação com linha de base ou estratégia capaz de enfrentar explicações alternativas. Para afirmar exploração, acrescentam-se conhecimento, intenção ou aproveitamento conforme a norma aplicável; não se infere dolo da existência do sistema.

Essa gradação permite que governança *ex ante* e responsabilidade *ex post* operem com padrões diferentes. Prevenção não precisa esperar causalidade individual consumada. Reparação exige conexão específica entre conduta, efeito e dano.

7.3 Programa de validação do protocolo

O protocolo ainda não foi validado. Um teste adequado deve: selecionar casos positivos e negativos; fornecer documentação equivalente a avaliadores independentes e cegos quanto à hipótese; elaborar manual de codificação; medir concordância entre avaliadores; testar validade discriminante contra aconselhamento legítimo, *default*, enquadramento e viés de automação; registrar falsos positivos e falsos negativos; e pré-registrar critérios de revisão.

Sua utilidade será enfraquecida se classificações dependerem predominantemente da convicção do avaliador, se casos legítimos forem sistematicamente rotulados como estruturais ou se conceitos existentes explicarem os mesmos fatos com igual precisão e menor complexidade.

8 CENÁRIOS DE APLICAÇÃO

8.1 Seleção e avaliação profissional

Um sistema classifica candidatos de 0 a 100 e exibe o escore antes do currículo. Recrutadores são avaliados por velocidade e aderência às recomendações. A empresa declara que “a decisão é humana”. A matriz revela domínio do ponto de partida, precedência,

consequência material e incentivo contra divergência. A independência não se prova pela possibilidade abstrata de alterar o escore.

Salvaguardas adequadas incluem análise preliminar de critérios essenciais antes da classificação, apresentação de evidências favoráveis e contrárias, justificativa curta para concordância em casos limítrofes, auditoria de taxas de divergência por grupo e proteção organizacional ao revisor. O registro não deve transformar o recrutador em autor exclusivo do erro do sistema.

8.2 Saúde

Uma ferramenta recomenda diagnóstico ou tratamento antes da avaliação médica e mostra uma única probabilidade em destaque. Em contexto de urgência e carga elevada, a recomendação pode fixar hipótese inicial e reduzir busca por sinais incompatíveis. A resposta não é esconder informação útil. É ordenar o fluxo: dados e alertas descritivos podem preceder a conclusão; diagnósticos alternativos e incerteza devem estar visíveis; discordâncias e desfechos precisam alimentar auditoria clínica.

Quando o sistema possui desempenho validado, sua influência pode melhorar cuidado. A governança deve comparar dois riscos: confiança excessiva e rejeição injustificada. O objetivo é calibração, não independência absoluta.

8.3 Crédito, preço e consumo

Uma plataforma conhece a faixa máxima que determinado usuário costuma aceitar e apresenta “preço anterior” ou parcela inicial personalizada. A referência pode ser legítima se verdadeira, comparável e explicada. Torna-se problemática quando não corresponde a prática comercial real, quando omite alternativas relevantes ou explora vulnerabilidade inferida. Logs de preço, critérios de personalização e testes de interface são mais informativos do que termos genéricos aceitos no cadastro.

Em crédito, um analista pode receber classificação de risco antes de examinar documentos e reproduzir o escore na justificativa final. A mera revisão humana não corrige dados errados ou discriminação indireta. É necessário registrar fontes, permitir correção, apresentar fatores incompatíveis e medir se decisões humanas apenas replicam a recomendação.

8.4 Justiça e administração pública

Um sistema sugere prioridade processual, faixa de acordo, minuta ou risco de reincidência. A referência pode aumentar consistência, mas também estreitar a consideração

do caso. Quanto maior a consequência sobre liberdade, renda ou acesso a direitos, maior a necessidade de juízo independente, razões individualizadas, rastreabilidade e possibilidade de impugnação.

Uma taxa de concordância de 98% não prova excelência. Pode refletir precisão do sistema, casos fáceis, pressão de produtividade ou ancoragem. A auditoria precisa comparar qualidade, divergências justificadas, erros evitados e distribuição por grupos. Concordância não é sinônimo de supervisão.

Quadro 7 — Cenários, risco principal e salvaguarda prioritária

Domínio	Âncora típica	Risco principal	Salvaguarda prioritária	Base jurídica possível
Trabalho	Escore ou recomendação de adequação	Exclusão e reprodução de vieses	Avaliação preliminar e auditoria de divergências	Proteção da personalidade, igualdade, LGPD e deveres laborais aplicáveis
Saúde	Diagnóstico, risco ou dose sugerida	Fechamento prematuro de hipótese	Evidência descritiva, alternativas e segunda verificação	Consentimento informado, dever de cuidado e proteção de dados sensíveis
Consumo	Preço anterior, desconto ou opção recomendada	Exploração de vulnerabilidade	Veracidade da referência e alternativas comparáveis	CDC, boa-fé, informação e práticas abusivas
Crédito	Classificação de risco	Recusa opaca e deferência humana	Correção de dados, fatores contrários e revisão efetiva	LGPD, CDC e regulação setorial
Judiciário	Faixa, ranking, minuta ou prognóstico	Ratificação e perda de individualização	Juízo preliminar, logs, motivação e contestação	Constituição, CPC e Resolução CNJ nº 615/2025
Administração	Elegibilidade ou prioridade sugerida	Automação material sem devido processo	Razões, canal de recurso e auditoria de impacto	Legalidade, motivação, devido processo e proteção de dados

Fonte: elaboração do autor.

9 SALVAGUARDAS: DA EXPLICAÇÃO À INDEPENDÊNCIA VERIFICÁVEL

9.1 A ordem da informação é variável de supervisão

Em decisões materiais, registrar avaliação inicial antes da recomendação pode criar linha de base e reduzir a influência de apoio incorreto. Buçinca, Malaya e Gajos (2021), Rastogi *et al.* (2022) e Agudo *et al.* (2024) oferecem suporte experimental a intervenções temporais e de forçamento cognitivo.

A salvaguarda não é universal. Kokje *et al.* (2026) mostram que aconselhamento não simultâneo pode reduzir sobreutilização e, ao mesmo tempo, aumentar rejeição de recomendações corretas, sem ganho global. A escolha entre apresentação preliminar, simultânea, sob demanda ou condicional deve considerar gravidade, reversibilidade, qualidade do sistema, experiência do usuário e risco de subutilização. O objetivo é calibrar dependência, não produzir independência absoluta.

9.2 Apresentação descritiva e pluralidade de hipóteses

Saídas prescritivas únicas concentram atenção. Conforme o domínio, podem-se apresentar faixas, incerteza, fatores favoráveis e contrários, casos semelhantes e alternativas plausíveis. Essa forma não elimina viés, mas distribui o campo comparativo e dificulta que um número se converta em veredito.

Uma lista de alternativas também pode ser manipulada por ordem, destaque ou exclusão. Pluralidade deve vir acompanhada de critérios de seleção e testes de compreensão. Mais informação não equivale automaticamente a melhor decisão.

9.3 Considerar o oposto e buscar evidência incompatível

O procedimento de considerar o oposto solicita razões pelas quais a recomendação pode estar errada. Carter e Liu (2025) encontraram efeito mitigador em seu contexto. A técnica é promissora, não universal. Pode tornar-se formalidade se reduzida a caixa obrigatória ou se a organização recompensar concordância.

Implementação robusta fornece acesso a evidências incompatíveis e protege divergência. O decisor pode indicar qual informação mudaria sua conclusão, produzindo critério refutável.

9.4 Explicabilidade não basta

Explicações podem aumentar compreensão e também autoridade. Cecil *et al.* (2024) e Vered *et al.* (2023) mostram que certas explicações não reduziram dependência excessiva. Três perguntas permanecem: houve compreensão; a explicação permitiu detectar erro; e existia possibilidade real de agir contra a saída?

9.5 Registros e métricas de independência

Logs úteis não armazenam apenas o clique final. Conforme risco e minimização de dados, preservam momento da exposição; versão de modelo e interface; referência; juízo preliminar, quando aplicável; alteração; razões de divergência; tempo; e resultado de recurso.

Taxa de concordância é ambígua. Deve ser combinada com qualidade, reversões, erros de comissão e omissão, distribuição por grupos e persistência depois da retirada. O objetivo não é maximizar divergência, mas verificar se ela é possível e produtiva.

Quadro 8 — Salvaguardas por etapa do ciclo decisório

Etapa	Salvaguarda	Evidência de implementação	Limite
Antes da exposição	Juízo preliminar em casos selecionados por risco.	Registro temporal e critérios mínimos.	Pode aumentar carga e subutilização da IA.

Etapa	Salvaguarda	Evidência de implementação	Limite
Apresentação	Faixas, incerteza, alternativas e fatores contrários.	Testes de interface e compreensão.	Excesso de informação pode sobrecarregar.
Deliberação	Considerar o oposto e buscar evidência incompatível.	Razões registradas e acesso aos dados.	Caixa formal não muda incentivos.
Decisão	Rejeição, reversão e segunda opinião efetivas.	Divergências e ausência de retaliação.	Opção técnica pode ser inviável na prática.
Pós-decisão	Recurso, correção e auditoria por grupos.	Decisões revistas, tempos e resultados.	Auditoria tardia não repara todos os danos.
Governança	Testes comparativos de sobreutilização e subutilização.	Protocolos, versões e relatórios independentes.	Métricas internas exigem verificação proporcional.

Fonte: elaboração do autor.

10 CONSEQUÊNCIAS JURÍDICAS E PROBATÓRIAS

10.1 Nenhum efeito jurídico automático

A categoria não cria vício do consentimento, responsabilidade, nulidade ou inversão automática do ônus. Sua função é qualificar fatos que podem preencher institutos existentes. Prevenção, exibição, redistribuição, valoração, invalidade e reparação possuem pressupostos distintos.

Quadro 9 — Respostas jurídicas possíveis e pressupostos próprios

Resposta possível	Base jurídica indicativa	Fato desencadeador	Prova e limite
Governança preventiva	LGPD, arts. 6º, 38, 46 e 50; regulação setorial.	Risco previsível e material do tratamento ou sistema.	Plausibilidade e proporcionalidade; não exige dano consumado.
Informação e revisão	CDC, arts. 6º e 46; LGPD, arts. 9º, 18 e 20.	Déficit informacional ou decisão abrangida pela norma.	Conteúdo, tempestividade e regime aplicável.
Preservação e exibição	CPC, arts. 369 e 396 a 400.	Evidência delimitada sob controle da outra parte.	Relevância, especificidade, sigilo e proteção de terceiros.
Distribuição dinâmica	CPC, art. 373, § 1º.	Impossibilidade ou maior facilidade concreta de produzir prova.	Decisão fundamentada, por questão, sem surpresa.
Redução da força probatória	CPC e valoração motivada.	Participação humana alegada, mas não documentada de modo verificável.	Avaliação do conjunto; ausência de log não prova falsidade por si.
Invalidade ou ineficácia	Código Civil, CDC ou regime específico.	Vício, abuso ou desconformidade legal comprovados.	Nexo entre arquitetura e requisito jurídico afetado.
Reparação	Regime de responsabilidade aplicável.	Conduta, dano, nexo e fator de imputação.	A matriz não substitui nenhum requisito.

Fonte: elaboração do autor. As bases são indicativas e dependem do caso.

10.2 Causalidade e contrafactual

Provar que alguém viu uma referência antes de decidir não basta. O desenho mais forte compara grupos expostos e não expostos, ordem de apresentação, versões de interface ou comportamento antes e depois da recomendação. Em litígio individual, podem ser utilizados

registros temporais, comparação entre equipes, padrões de deslocamento e persistência, sempre com hipóteses rivais.

Taxa elevada de concordância é ambígua. Pode refletir precisão, casos fáceis, pressão organizacional ou ancoragem. Textos que reproduzem a recomendação constituem indício, não prova conclusiva. A análise precisa enfrentar causalidade múltipla e explicações alternativas.

10.3 Assimetria de acesso à prova

O afetado geralmente não possui versão da interface, dados de segmentação ou histórico de experimentos. Exigir prova do mecanismo interno antes da exibição tornaria o direito circular. Revelação integral, porém, pode afetar segredos, segurança e dados de terceiros. A resposta deve ser delimitada: preservação por questão, acesso controlado, perícia, anonimização e ordem específica.

A maior facilidade de produzir prova pode ser considerada nos termos do CPC, mediante decisão fundamentada e oportunidade de cumprimento (Didier Jr.; Braga; Oliveira, 2026). Redistribuição não decorre do rótulo “IA”, mas do controle demonstrado sobre evidência relevante e da dificuldade concreta da outra parte.

10.4 Decisão humana como questão de prova

Quando o fornecedor afirma que o sistema apenas auxiliou, a participação humana deve ser verificável. Indícios de atuação significativa incluem juízo anterior quando proporcional, informação independente, competência, tempo, poder de rejeição, divergências reais e razões próprias. Indícios de ratificação incluem recomendação exibida primeiro, metas de adesão, justificativas copiadas, baixa divergência após erros e punição por afastamento.

A distinção não elimina responsabilidade humana. Distribui-a conforme controle efetivo. Uma organização não deve utilizar presença formal do trabalhador para absorver o risco criado por modelo, interface e incentivos que definiu.

11 OBJEÇÕES, LIMITES E FALSIFICABILIDADE

11.1 “Toda decisão possui um ponto de partida”

A objeção é correta. Não existe deliberação sem contexto ou referência. O conceito identifica quando o controle se torna persistente ou repetível, material e pouco contestável. Âncoras públicas, transparentes e revisáveis podem aumentar igualdade. O problema é sua governança.

11.2 “A IA pode ser mais precisa do que o humano”

Também é correto. Recomendação validada pode reduzir erro. Autonomia não exige ignorar bons conselhos; exige dependência calibrada, limites conhecidos e reconhecimento de exceções. Precisão média não resolve distribuição de erros, mudança de contexto ou contestação.

11.3 “O conceito é amplo demais”

O risco de expansão é real. Por isso, a qualificação exige os três núcleos cumulativos do Quadro 6. Sem controle relevante, capacidade de deslocamento material ou déficit proporcional de independência, deve-se permanecer nas categorias de aconselhamento, design, viés de automação ou ancoragem episódica.

11.4 Hipóteses de refutação

A proposta será enfraquecida se pesquisas demonstrarem consistentemente que: a ordem de apresentação não altera decisões nos contextos examinados; personalização e repetição não ampliam nem estabilizam efeitos; salvaguardas não produzem diferença útil; ou supervisores com poder de rejeição apresentam os mesmos padrões daqueles sujeitos a metas de adesão.

Também exigirá revisão se avaliadores independentes não alcançarem concordância aceitável; se o protocolo não discriminar aconselhamento legítimo de exploração estrutural; se produzir falsos positivos sistemáticos; ou se conceitos existentes explicarem os casos com igual precisão e menor complexidade.

11.5 Agenda empírica brasileira

Uma agenda de validação deve combinar:

1. experimentos pré-registrados: comparar recomendação antes, simultaneamente e depois do juízo humano, medindo sobreutilização e subutilização;

2. codificação cega de casos: aplicar manual do protocolo por avaliadores independentes e medir concordância;

3. estudos de campo: analisar divergência, reversão e erro em organizações que adotam apoio algorítmico;

4. jurisprudência: mapear decisões automatizadas ou assistidas, prova exigida e tratamento da participação humana;

5. auditorias de interface: documentar referências, personalização, ordem e compreensão em consumo, crédito, trabalho e serviços públicos.

Os estudos devem definir linhas de base, medir efeitos por grupos, registrar falsos positivos e distinguir precisão de conformidade. A ausência de divergência não pode ser usada isoladamente como indicador de qualidade.

12 CONCLUSÃO

A decisão visível não é necessariamente o início da decisão. Em ambientes algorítmicos, referências podem ser escolhidas antes que o sujeito examine alternativas, adaptadas a seu histórico, repetidas e avaliadas por sua capacidade de produzir adesão. O fenômeno clássico da ancoragem adquire dimensão institucional quando o controle do ponto de partida é persistente ou repetível e a independência se torna proporcionalmente insuficiente.

O artigo não atribui intenção à tecnologia, não presume manipulação e não reivindica novidade para ancoragem algorítmica, juízo humano preliminar ou intervenções de desancoramento. Sua contribuição incremental é jurídico-operacional: separar mecanismo cognitivo, qualificação estrutural e consequência jurídica; reconstruir essa relação no direito brasileiro; e propor protocolo heurístico não pontuado de oito eixos e três núcleos cumulativos.

O protocolo não produz efeitos jurídicos. Prevenção, informação, exibição, distribuição dinâmica, valoração, invalidade e reparação dependem de bases normativas e pressupostos próprios. A presença de um humano ao final do fluxo, por sua vez, não prova independência; deve ser examinada por competência, tempo, informação alternativa, poder de rejeição e divergências documentadas.

As salvaguardas precisam ser contextuais. Juízo preliminar e aconselhamento não simultâneo podem reduzir sobreutilização e aumentar subutilização. A pergunta adequada não é como eliminar influência, mas como calibrar dependência, registrar cronologia e preservar contestação sem desperdiçar recomendações corretas.

O limite principal permanece aberto: o protocolo ainda não foi validado em amostra brasileira nem submetido a concordância entre avaliadores. Sua força atual é dogmática e heurística. Seu futuro dependerá da capacidade de distinguir casos legítimos e problemáticos, resistir a falsos positivos e produzir perguntas probatórias mais precisas do que as categorias já disponíveis.

O desafio final não é impedir que decisões humanas tenham referências. É impedir que quem controla silenciosamente o ponto de partida possa medir e ajustar seus efeitos enquanto

transfere ao indivíduo toda a responsabilidade pelo resultado. Quando isso ocorre, o clique permanece humano; a arquitetura anterior precisa tornar-se juridicamente visível.

DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Utilizou-se inteligência artificial generativa como apoio instrumental na organização bibliográfica, revisão linguística e formatação. A delimitação do problema, a construção da tese, a seleção crítica das fontes, os argumentos e a aprovação final são de responsabilidade exclusiva do autor.

REFERÊNCIAS

ADAM, Hammaad *et al.* Mitigating the impact of biased artificial intelligence in emergency decision-making. *Communications Medicine*, v. 2, art. 149, 2022. DOI: 10.1038/s43856-022-00214-4. Disponível em: https://doi.org/10.1038/s43856-022-00214-4. Acesso em: 2 ago. 2026.

AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS (ANPD). *Resultados da Tomada de Subsídios sobre Inteligência Artificial e Revisão de Decisões Automatizadas*. Brasília, DF, 19 maio 2025. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-apresenta-resultados-da-tomada-de-subsidios-sobre-tratamento-automatizado-de-dados-pessoais/. Acesso em: 2 ago. 2026.

AGUDO, Ujué *et al.* The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, v. 9, art. 1, 2024. DOI: 10.1186/s41235-023-00529-3. Disponível em: https://doi.org/10.1186/s41235-023-00529-3. Acesso em: 2 ago. 2026.

BIONI, Bruno Ricardo. *Proteção de dados pessoais: a função e os limites do consentimento*. 3. ed. Rio de Janeiro: Forense, 2021. 344 p. ISBN 978-85-309-9408-2.

BRASIL. *Lei nº 10.406, de 10 de janeiro de 2002*. Institui o Código Civil. Brasília, DF: Presidência da República, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em: 2 ago. 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/l13105compilada.htm. Acesso em: 2 ago. 2026.

BRASIL. *Lei nº 13.709, de 14 de agosto de 2018*. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709compilado.htm. Acesso em: 2 ago. 2026.

BRASIL. *Lei nº 8.078, de 11 de setembro de 1990*. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 2 ago. 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. Alterada pela Resolução nº 674/2026. Disponível em: https://atos.cnj.jus.br/atos/detalhar/6001. Acesso em: 2 ago. 2026.

BRASIL. [Constituição (1988)]. *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: 2 ago. 2026.

BUÇINCA, Zana; MALAYA, Maja Barbara; GAJOS, Krzysztof Z. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, v. 5, n. CSCW1, art. 55, p. 1-21, 2021. DOI: 10.1145/3449287. Disponível em: https://doi.org/10.1145/3449287. Acesso em: 2 ago. 2026.

CALO, Ryan. Digital market manipulation. *George Washington Law Review*, v. 82, p. 995-1051, 2014. Disponível em: https://digitalcommons.law.uw.edu/faculty-articles/25/. Acesso em: 2 ago. 2026.

CARTER, Lemuria; LIU, Dapeng. How was my performance? Exploring the role of anchoring bias in AI-assisted decision making. *International Journal of Information Management*, v. 82, art. 102875, 2025. DOI: 10.1016/j.ijinfomgt.2025.102875. Disponível em: https://doi.org/10.1016/j.ijinfomgt.2025.102875. Acesso em: 2 ago. 2026.

CECIL, Julia *et al.* Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task. *Scientific Reports*, v. 14, art. 9736, 2024. DOI: 10.1038/s41598-024-60220-5. Disponível em: https://doi.org/10.1038/s41598-024-60220-5. Acesso em: 2 ago. 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.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. 3. ed. São Paulo: Thomson Reuters Brasil, 2021. 368 p. ISBN 978-65-5991-796-9.

ENGLICH, Birte; MUSSWEILER, Thomas; STRACK, Fritz. Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, v. 32, n. 2, p. 188-200, 2006. DOI: 10.1177/0146167205282152. Disponível em: https://doi.org/10.1177/0146167205282152. Acesso em: 2 ago. 2026.

FURNHAM, Adrian; BOO, Hua Chu. A literature review of the anchoring effect. *The Journal of Socio-Economics*, v. 40, n. 1, p. 35-42, 2011. DOI: 10.1016/j.socec.2010.10.008. Disponível em: https://doi.org/10.1016/j.socec.2010.10.008. Acesso em: 2 ago. 2026.

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.

GLICKMAN, Moshe; SHAROT, Tali. How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, v. 9, p. 345-359, 2025. DOI: 10.1038/s41562-024-02077-2. Disponível em: https://doi.org/10.1038/s41562-024-02077-2. Acesso em: 2 ago. 2026.

HÜTTER, Mandy; FIEDLER, Klaus. Advice taking under uncertainty: the impact of genuine advice versus arbitrary anchors on judgment. *Journal of Experimental Social Psychology*, v. 85, art. 103829, 2019. DOI: 10.1016/j.jesp.2019.103829. Disponível em: https://doi.org/10.1016/j.jesp.2019.103829. Acesso em: 2 ago. 2026.

IKEDA, Shinnosuke. Inconsistent advice by ChatGPT influences decision making in various areas. *Scientific Reports*, v. 14, art. 15876, 2024. DOI: 10.1038/s41598-024-66821-4. Disponível em: https://doi.org/10.1038/s41598-024-66821-4. Acesso em: 2 ago. 2026.

JAIN, Gaurav; NAYAKANKUPPAM, Dhananjay; GAETH, Gary J. Perceptual anchoring and adjustment. *Journal of Behavioral Decision Making*, v. 34, n. 4, p. 581-592, 2021. DOI: 10.1002/bdm.2231. Disponível em: https://doi.org/10.1002/bdm.2231. Acesso em: 2 ago. 2026.

KÖCHER, Sören *et al.* New hidden persuaders: an investigation of attribute-level anchoring effects of product recommendations. *Journal of Retailing*, v. 95, n. 1, p. 24-41, 2019. DOI: 10.1016/j.jretai.2018.10.004. Disponível em: https://doi.org/10.1016/j.jretai.2018.10.004. Acesso em: 2 ago. 2026.

KOKJE, Eesha *et al.* AI-augmented decision-making in face matching: comparing concurrent and non-concurrent advice presentation. *Cognitive Research: Principles and Implications*, v. 11, art. 11, 2026. DOI: 10.1186/s41235-026-00707-z. Disponível em: https://doi.org/10.1186/s41235-026-00707-z. Acesso em: 2 ago. 2026.

KOULU, Riikka. Proceduralizing control and discretion: human oversight in artificial intelligence policy. *Maastricht Journal of European and Comparative Law*, v. 27, n. 6, p. 720-735, 2020. DOI: 10.1177/1023263X20978649. Disponível em: https://doi.org/10.1177/1023263X20978649. Acesso em: 2 ago. 2026.

LAUX, Johann; RUSCHEMEIER, Hannah. Automation bias in the AI Act: on the legal implications of attempting to de-bias human oversight of AI. *European Journal of Risk Regulation*, v. 16, n. 4, p. 1519-1534, 2025. DOI: 10.1017/err.2025.10033. Disponível em: https://doi.org/10.1017/err.2025.10033. Acesso em: 2 ago. 2026.

MARQUES, Claudia Lima. *Contratos no Código de Defesa do Consumidor*. 10. ed. São Paulo: Thomson Reuters Brasil, 2025. 1680 p. ISBN 978-65-260-1400-4.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação algorítmica: conceito, fundamento legal e tipologia. *Direito Público*, v. 16, n. 90, p. 39-64, 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766. Acesso em: 2 ago. 2026.

MIRAGEM, Bruno. *Curso de direito do consumidor*. 9. ed. Rio de Janeiro: Forense, 2024. 1304 p. ISBN 978-65-5964-884-9.

MOSIER, Kathleen L.; SKITKA, Linda J. Automation use and automation bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, v. 43, n. 3, p. 344-348, 1999. DOI: 10.1177/154193129904300346. Disponível em: https://doi.org/10.1177/154193129904300346. Acesso em: 2 ago. 2026.

NARAYANAN, Pranadharthiharan; SOMASUNDARAM, Jeeva; SEIFERT, Matthias. Risk-averse algorithmic support and inventory management. *European Journal of Operational Research*, v. 322, n. 3, p. 993-1004, 2025. DOI: 10.1016/j.ejor.2024.11.013. Disponível em: https://doi.org/10.1016/j.ejor.2024.11.013. Acesso em: 2 ago. 2026.

NGUYEN, Jeremy K. Human bias in AI models? Anchoring effects and mitigation strategies in large language models. *Journal of Behavioral and Experimental Finance*, v. 43, art. 100971, 2024. DOI: 10.1016/j.jbef.2024.100971. Disponível em: https://doi.org/10.1016/j.jbef.2024.100971. Acesso em: 2 ago. 2026.

PARASURAMAN, Raja; MANZEY, Dietrich H. Complacency and bias in human use of automation: an attentional integration. *Human Factors*, v. 52, n. 3, p. 381-410, 2010. DOI: 10.1177/0018720810376055. Disponível em: https://doi.org/10.1177/0018720810376055. Acesso em: 2 ago. 2026.

RASTOGI, Charvi *et al.* Deciding fast and slow: the role of cognitive biases in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, v. 6, n. CSCW1, art. 83, p. 1-22, 2022. DOI: 10.1145/3512930. Disponível em: https://doi.org/10.1145/3512930. Acesso em: 2 ago. 2026.

SCHMAUDER, Christian *et al.* Algorithmic nudging: the need for an interdisciplinary oversight. *Topoi*, v. 42, n. 3, p. 799-807, 2023. DOI: 10.1007/s11245-023-09907-4. Disponível em: https://doi.org/10.1007/s11245-023-09907-4. Acesso em: 2 ago. 2026.

SCHREIBER, Anderson. *Manual de direito civil contemporâneo*. 9. ed. São Paulo: SaraivaJur, 2026. 848 p. ISBN 978-65-5177-080-7.

STRIKAITĖ-LATUŠINSKAJA, Goda. Balancing legal certainty and judicial discretion in AI-assisted adjudication. In: ZALNIERIUTE, Monika; LIMANTE, Agne (ed.). *The Cambridge handbook of AI and technologies in courts*. Cambridge: Cambridge University Press, 2026. p. 293-306. DOI: 10.1017/9781009744225.024. Disponível em: https://doi.org/10.1017/9781009744225.024. Acesso em: 2 ago. 2026.

SUSSER, Daniel; ROESSLER, Beate; NISSENBAUM, Helen. Online manipulation: hidden influences in a digital world. *Georgetown Law Technology Review*, v. 4, n. 1, p. 1-45, 2019. Disponível em: https://georgetownlawtechreview.org/online-manipulation-hidden-influences-in-a-digital-world/GLTR-01-2020/. Acesso em: 2 ago. 2026.

TEICHMAN, Doron; ZAMIR, Eyal; RITOV, Ilana. Biases in legal decision-making: comparing prosecutors, defense attorneys, law students, and laypersons. *Journal of Empirical Legal Studies*, v. 20, n. 4, p. 852-894, 2023. DOI: 10.1111/jels.12365. Disponível em: https://doi.org/10.1111/jels.12365. Acesso em: 2 ago. 2026.

TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty: heuristics and biases. *Science*, v. 185, n. 4157, p. 1124-1131, 1974. DOI: 10.1126/science.185.4157.1124. Disponível em: https://doi.org/10.1126/science.185.4157.1124. Acesso em: 2 ago. 2026.

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. *Jornal Oficial da União Europeia*, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng. Acesso em: 2 ago. 2026.

VERED, Mor *et al.* The effects of explanations on automation bias. *Artificial Intelligence*, v. 322, art. 103952, 2023. DOI: 10.1016/j.artint.2023.103952. Disponível em: https://doi.org/10.1016/j.artint.2023.103952. Acesso em: 2 ago. 2026.

YEUNG, Karen. "Hypernudge": big data as a mode of regulation by design. *Information, Communication & Society*, v. 20, n. 1, p. 118-136, 2017. DOI: 10.1080/1369118X.2016.1186713. Disponível em: https://doi.org/10.1080/1369118X.2016.1186713. Acesso em: 2 ago. 2026.