

RELATÓRIO TÉCNICO DA RODADA EXPERIMENTAL

Comparação metodológica entre benchmark humano preexistente e quatro execuções de IA
VERSÃO COMPLETA ANONIMIZADA PARA PUBLICAÇÃO

Campo	Identificação
Processo	Identificação suprimida na versão pública
Incidente	Desconsideração inversa da personalidade jurídica em cumprimento de sentença
Benchmark humano	Impugnação preexistente da pessoa jurídica terceira, protocolada em abril de 2025
Decisão posterior principal	Decisão posterior, proferida em julho de 2025, que determinou a instauração do IDPJ em autos apartados
Condição MVP	Execução 1, conforme o mapeamento adotado na fase comparativa desta rodada
Data da consolidação	26/08/2026

Nota metodológica e de publicação. Este relatório registra uma rodada experimental realizada sobre corpus processual histórico e anonimizado proveniente de caso real. A comparação tem finalidade técnica e metodológica. Os resultados são restritos a esta rodada e não representam promessa de desempenho, validação do sistema ou afirmação de superioridade geral em relação ao trabalho humano. A decisão judicial posterior é tratada apenas como evidência ex post sobre o fundamento efetivamente adotado, e não como padrão-ouro retroativo de correção.

1. Síntese executiva da rodada

Resultado central. Na auditoria cega das quatro execuções de IA, a Execução 1 (MVP) apresentou a maior aderência global ao conjunto de critérios avaliados depois que sua minuta final foi incorporada ao corpus. Esse resultado descreve somente as condições testadas nesta rodada e não autoriza inferência geral sobre pessoas, ferramentas ou prática jurídica. A distribuição variou por dimensão: o benchmark humano se alinhou de forma mais direta ao ponto processual posteriormente adotado pelo juízo; o MVP recuperou fatos adversos não explorados no benchmark, estruturou o teste adversarial e preservou de forma mais explícita a distinção entre prova e inferência. [S2]

Achados registrados no benchmark humano. A impugnação preexistente identificou de forma expressa a necessidade de instauração regular do incidente, atacou o uso do art. 134, §2º, do CPC para justificar o processamento adotado e pediu solução processual coincidente com a decisão posterior de autuação apartada. Também explorou diretamente a distinção entre empresário individual e sociedade limitada unipessoal e formulou pedido de honorários. [S1]

Achados registrados na condição MVP. O MVP identificou como principal risco adverso a constituição da sociedade durante uma execução antiga, utilizou o histórico de um bem anteriormente localizado para qualificar a narrativa de frustração executiva absoluta, recomendou prova positiva de autonomia patrimonial e evitou converter ausência de prova adversária em prova de inexistência de confusão patrimonial. [S2]

Achado exclusivo relevante de outra execução. A Execução 3 foi a única a detectar divergência entre o endereço usado no requerimento e no despacho e o endereço constante do cadastro oficial da pessoa jurídica. Esse achado não apareceu no benchmark humano, no MVP nem nas Execuções 2 e 4. A mesma execução, contudo, apresentou falha fundacional: presumiu o cliente e produziu peça para o polo oposto. [S2]

Limitação crítica de comparabilidade histórica. As Execuções 1 e 4 utilizaram, segundo a auditoria cega, precedentes do STJ julgados em 2026, posteriores ao momento histórico do benchmark humano, produzido em abril de 2025. Esses fundamentos refletem assimetria de acesso temporal e foram excluídos da comparação de desempenho ex ante. [S2]

2. Identificação do processo e contexto jurídico

O caso envolve cumprimento de sentença entre partes mantidas sob anonimato. Em setembro de 2024, o exequente requereu desconsideração inversa para alcançar o patrimônio de uma sociedade empresária limitada unipessoal, da qual o devedor originário figurava como sócio-administrador único. O pedido partiu, em essência, da longa duração da execução, da alegada frustração de medidas constritivas e da inferência de que a unicidade societária tornaria 'evidente' a confusão patrimonial. [S1]

A documentação oficial juntada com o pedido identificava a pessoa jurídica terceira como Sociedade Empresária Limitada, aberta em outubro de 2022, ativa e administrada pelo devedor originário. O precedente transcrito pelo exequente envolvia empresário individual, figura juridicamente distinta. O histórico da execução também continha restrição e avaliação de veículo, além de diligências de busca e apreensão frustradas por ausência de meios de remoção e guarda atribuída à parte exequente. [S1/S2]

Em fevereiro de 2025, o juízo determinou a citação da pessoa jurídica terceira, suspendeu o cumprimento de sentença e previu manifestação e produção de prova. O benchmark humano foi protocolado em abril de 2025. Em julho de 2025, o juízo tornou sem efeito o despacho anterior e determinou a instauração do incidente em autos apartados, com prosseguimento da execução contra o devedor originário. O mérito do art. 50 do Código Civil não foi decidido nessa ocasião. [S1]

3. Corpus e sequência efetivamente utilizados

Ref.	Elemento	Função experimental
C1	Corpus fático cego	Extrato processual disponibilizado às execuções de IA, com corte anterior à benchmark humano e sem a decisão posterior.
C2	Quatro execuções de IA	Respostas independentes, auditadas isoladamente antes da comparação transversal.
C3	Minuta final da Execução 1	Incorporada posteriormente ao corpus da auditoria cega, eliminando a limitação de auditabilidade fim a fim da condição MVP.
C4	Benchmark humano preexistente	Impugnação real, auditada ex ante apenas com o corpus historicamente disponível.
C5	Decisão judicial posterior	Decisão proferida em julho de 2025, examinada em bloco separado, sem retroalimentar a avaliação ex ante do benchmark humano.
C6	Exportação processual posterior completa	Usada nesta consolidação para confirmar a sequência real dos atos e decisões e para identificar limites do corpus cego original.

Observação temporal importante: a auditoria cega registrou que o PDF fornecido às IAs não comprovava de forma segura a citação efetiva. A exportação processual posterior, mantida em arquivo reservado, contém certidão de citação. Portanto, a classificação da Execução 2 como erro por escrever 'citada' permanece válida em relação ao input experimental, embora o ato tenha ocorrido na realidade. O erro foi de integridade input-output, não uma afirmação de que a citação jamais aconteceu.

4. Metodologia efetivamente utilizada

Auditoria cega individual: cada execução foi confrontada primeiro com o corpus disponível, sem ranking prévio e sem uso do benchmark humano como gabarito.

Crerios: fidelidade documental; identificação de cliente, posição e fase; reconstrução da controvérsia; distinção entre alegação, prova e inferência; omissões; contradições; riscos; teses; melhor argumento adverso; estratégia; medida processual; integridade análise-redação; necessidade de correção humana.

Classificação das divergências: redacional; interpretativa/estratégica defensável; omissão material; erro

Comparação final: somente após as auditorias separadas foram cruzados humano, MVP e demais execuções. [S2]

5. As quatro condições experimentais [S2]

Condição	Comportamento observado	Aspecto registrado	Falha/limite principal
Execução 1 - MVP	Bloqueou redação até conhecer o cliente; defendeu a pessoa jurídica; reconstruiu fatos, cronologia e risco adverso; recomendou prova positiva; minuta final auditável.	Maior aderência global entre as condições de IA nesta rodada.	Omissão do conflito de endereço; pequena imprecisão descritiva de precedente; uso de autoridade posterior ao corte histórico.
Execução 2	Bloqueou até identificar o cliente; confrontou o precedente adverso e recuperou a cronologia e o histórico do bem.	Confrontou o mérito e o precedente com apoio documental.	Transformou incerteza sobre citação em fato certo na minuta; não detectou o conflito de endereço; menor ênfase em prova positiva.
Execução 3	Reconstruiu o estado processual e detectou a divergência de endereço.	Única condição a registrar a divergência de endereço.	Presumiu o cliente como exequente e redigiu para o polo oposto; erro de medida processual e de objetivo.
Execução 4	Bloqueou enquanto o cliente era desconhecido; explicitou cautela probatória; tratou a citação com reserva; recomendou prova positiva.	Aderência global próxima à Execução 1; distinção prova-inferência especialmente explícita.	Omitiu o principal fato adverso cronológico; não detectou o conflito de endereço; uso de autoridade posterior ao corte histórico.

6. Resultado da auditoria cega das IAs

A classificação interna da auditoria cega, após incorporação da minuta da Execução 1, registrou: Execução 1 com maior aderência global ao conjunto de critérios; Execução 4 com aderência próxima e distinção especialmente explícita entre prova e inferência; Execução 2 com aderência no confronto de mérito, mas comprometida por erro factual/documental na minuta; e Execução 3 comprometida por presunção indevida do cliente e produção de peça para o polo errado. [S2]

Essa classificação descreve as quatro condições sob o desenho realizado. Ela não deve ser convertida em ranking geral entre trabalho humano e IA, porque: (i) o benchmark humano foi produzido em contexto real, com limitações e objetivos próprios; (ii) o resultado judicial posterior não é gabarito de todas as teses; e (iii) parte da fundamentação jurídica das IAs utilizou jurisprudência posterior ao momento histórico do benchmark.

7. Avaliação independente do benchmark humano

A avaliação independente do benchmark humano registrou aderência relevante em: identificação correta da empresa como terceira atingida; distinção entre empresário individual e sociedade limitada unipessoal; defesa da autonomia patrimonial; exigência dos pressupostos do art. 50 do Código Civil; contestação da ideia de que a mera falta de bens bastaria; percepção do problema procedimental do IDPJ; e pedido de honorários. [S1]

As principais fragilidades registradas foram: ausência de uso do histórico concreto do bem já constrito e avaliado; não enfrentamento da constituição da sociedade em 2022 durante execução antiga; ausência de prova positiva contábil e bancária de separação patrimonial; afirmações mais categóricas do que a prova permitia; tese excessiva de que a desconsideração inversa dependeria de transferência formal de bens; referências normativas imprecisas ('49-A e 50 do CPC' e menção a 'CPP'); e pedidos finais que não espelharam com precisão toda a preliminar processual desenvolvida. [S1]

8. Confronto por dimensões entre benchmark humano, MVP e demais execuções

8.1. Tipologia usada nesta comparação

Categoria	Definição operacional
Diferença material observada	Diferença que altera de forma substantiva a identificação de fatos, riscos, prova, medida ou tese e apresenta utilidade prática no caso.
Diferença estratégica defensável	Escolha distinta, mas juridicamente e operacionalmente sustentável; não é erro nem sinal automático de maior qualidade.
Diferença apenas redacional	Variação de estrutura, extensão, nomenclatura ou densidade sem efeito relevante sobre risco ou correção.
Omissão	Fato, documento, risco ou pedido relevante disponível no corpus e não utilizado.
Erro	Afirmação ou medida incompatível com o corpus, o cliente, a fase ou a disciplina jurídica aplicável.

8.2. Matriz transversal

Dimensão	Benchmark humano	MVP/E1	E2	E3	E4	Descrição comparativa
Cliente, polo e objetivo	Correto	Correto	Correto	Erro crítico	Correto	MVP e benchmark humano identificaram corretamente cliente e objetivo; E3 apresentou erro de polo.
Fase e medida defensiva	Correto; priorizou o IDPJ	Correto	Correto	Peça para polo oposto	Correto	O benchmark humano priorizou o ponto processual adotado depois.
Ltda. x empresário individual	Explícito e detalhado	Correto, menos explorado	Confronto documental detalhado	Não desenvolve defesa	Correto por outra via	O benchmark humano e E2 exploraram diretamente a distinção documental.
Constituição durante execução	Omitiu	Identificou e enfrentou	Identificou e enfrentou	Não chega à defesa	Omitiu	MVP e E2 enfrentaram o fato cronológico; ele não foi explorado no benchmark humano nem em E4.
Histórico de bem localizado	Omitiu	Usou com cautela	Usou	Não desenvolveu	Usou	MVP, E2 e E4 recuperaram o histórico; o benchmark humano não o utilizou.
Prova positiva de autonomia	Insuficiente	Recomendou	Reativo	Não aplicável	Recomendou e explicitou a cautela	MVP e E4 indicaram prova positiva; o benchmark não.
Prova x inferência	Algumas conclusões categóricas	Separou prova e inferência	Separou na análise; vazou na minuta	Presunção de cliente	Distinção explícita entre prova e inferência	MVP e E4 preservaram a distinção; E2 perdeu integridade na passagem para a minuta.
Endereço divergente	Não explorou	Omitiu	Omitiu	Detectou	Omitiu	E3 foi a única condição a registrar esse achado.
Honorários	Pediu	Pediu	Não pediu	Não central	Pediu	Diferença estratégica defensável; não decisiva.
Teste adversarial	Não enfrentou o principal fato	Enfrentou	Enfrentou	Comprometido pelo polo errado	Incompleto por omitir cronologia	MVP e E2 enfrentaram o principal fato adverso; o benchmark humano não o explorou.
Proximidade com decisão posterior	Alta no núcleo procedimental	Moderada: IDPJ correto, menor ênfase no autos apartados	Moderada	Baixa para o interesse da pessoa jurídica terceira	Moderada	O benchmark humano se alinhou de forma mais direta ao fundamento efetivamente usado.

8.3. Convergências entre o benchmark humano e o MVP

- Identificação da pessoa jurídica terceira como pessoa jurídica diretamente atingida pelo pedido e necessidade de defesa no âmbito do IDPJ.
- Autonomia patrimonial da sociedade limitada e rejeição da ideia de confusão automática pela unicidade societária.
- Centralidade do art. 50 do Código Civil e necessidade de fatos concretos de abuso/confusão, não apenas insolvência.
- Cabimento de manifestação defensiva com produção de prova.

- Utilidade de preservar pedido de honorários, embora ele não tenha sido determinante para a decisão posterior.

8.4. Aspectos mais desenvolvidos no benchmark humano

- O benchmark humano atacou frontalmente o fundamento processual usado pelo despacho anterior e sustentou a necessidade de instauração formal do incidente, com ênfase no processamento apartado. Esse ponto coincidiu com o núcleo posteriormente acolhido pelo juízo.
- A distinção textual e retórica entre empresário individual e sociedade limitada unipessoal foi explorada diretamente pelo benchmark humano. O MVP sustentou a litude da unipessoalidade, mas a auditoria cega registra que não explicitou a diferença entre o precedente adversário de empresário individual e a natureza limitada da pessoa jurídica terceira.
- O benchmark humano desenvolveu de forma extensa a tese de honorários e causalidade. O MVP também preservou o pedido, mas o tema não foi determinante naquele momento e, por isso, não serve como eixo material de comparação.

8.5. Aspectos identificados pelo MVP e não utilizados no benchmark humano

- Abertura da sociedade durante execução antiga como principal argumento adverso de persuasão, com tratamento explícito de que a cronologia gera risco, mas não prova fraude.
- Histórico do bem já localizado, avaliado e objeto de diligências frustradas por falta de meios de remoção/guarda atribuída ao exequente, útil para qualificar a narrativa de 'todas as constrições infrutíferas'.
- Necessidade de prova positiva de autonomia patrimonial - documentos contábeis, bancários, fiscais e societários - antes de afirmar separação patrimonial perfeita.
- Disciplina para separar 'o requerente não provou confusão patrimonial' de 'não existe confusão patrimonial'.
- Teste adversarial estruturado em torno do pior fato para a defesa, em vez de apenas responder à formulação mais fraca do exequente.

8.6. Achados exclusivos ou mais explícitos nas demais execuções

- Execução 3: divergência interdocumental de endereço, potencialmente capaz de afetar a citação. Foi um achado material e exclusivo.
- Execução 2: confrontou a qualificação 'ME' como porte com a natureza jurídica limitada e registrou a inadequação do precedente adversário sobre empresário individual. O mesmo ponto existia no benchmark humano; a Execução 2 explicitou com maior detalhamento a incompatibilidade documental.
- Execução 4: formulou de modo mais explícito a regra de não afirmar 'separação patrimonial perfeita' sem documentos internos. O MVP também recomendou prova positiva; a diferença observada esteve no grau de explicitação dessa cautela.

9. Erros factuais, documentais e jurídicos

Produto	Erro/omissão	Gravidade
Benchmark humano	Imprecisões normativas ('49-A e 50 do CPC'; referência a 'CPP'); afirmação excessiva de que a desconsideração inversa exigiria transferência formal; extrapolação de autonomia jurídica para inexistência factual de confusão.	Material, mas não fundacional.
MVP / E1	Omissão do conflito de endereço; pequena imprecisão descritiva sobre um precedente; expressão 'dado neutro' para a cronologia foi excessiva. Parte da jurisprudência usada é posterior ao momento histórico do benchmark humano.	Localizado; exige ajuste e controle temporal.
Execução 2	A análise reconheceu incerteza sobre citação, mas a minuta afirmou que a empresa estava 'citada'.	Crítico no experimento: quebra de integridade análise-redação.
Execução 3	Presumiu o cliente sem base e redigiu para o exequente, embora o objetivo efetivo fosse defender a terceira atingida.	Crítico/fundacional; exige refazer a execução.
Execução 4	Omitiu o principal fato adverso concreto: constituição da sociedade durante execução antiga; também não detectou conflito de endereço. Parte da jurisprudência usada é posterior ao corte histórico.	Material, mas estrutura aproveitável.

10. Qualidade do teste adversarial, estratégia e disciplina probatória

Teste adversarial. O MVP e a Execução 2 enfrentaram explicitamente o fato que um adversário competente poderia explorar: sociedade constituída durante execução prolongada, sob controle exclusivo do devedor originário. O benchmark humano atacou a tese fraca ('sócio único = confusão'), mas não neutralizou o argumento adverso mais consistente ('cronologia + controle + execução antiga justificam investigação'). A Execução 4 não registrou esse fato, o que reduziu sua aderência nesta dimensão.

Estratégia. O benchmark humano priorizou uma objeção processual capaz de produzir resultado imediato sem exigir julgamento do mérito patrimonial. O MVP estruturou defesa mais extensa para eventual julgamento de mérito, mas não registrou, no relatório cego, a mesma ênfase no ponto específico de autuação apartada. As abordagens são complementares: uma peça operacionalmente mais completa reuniria a preliminar processual do benchmark humano e o pacote probatório e adversarial do MVP.

Disciplina prova-inferência. O MVP e a Execução 4 distinguiram fatos documentados, ausência de prova do adversário e inexistência objetiva de confusão; a Execução 4 explicitou essa cautela com maior extensão. O benchmark humano, embora tenha rejeitado corretamente a presunção automática, formulou em alguns trechos a inexistência de confusão com grau de certeza superior ao suporte documental disponível.

11. Relação entre argumentos e decisão judicial posterior

A decisão de julho de 2025 não decidiu se havia abuso, desvio de finalidade ou confusão patrimonial. Seu núcleo foi processual: terceiro estranho à relação originária, necessidade de procedimento próprio, contraditório e citação, com autuação do incidente em separado e prosseguimento da execução contra o devedor original. [S1]

Fundamento	Quem apresentou	Tratamento judicial	Peso no resultado
Necessidade de IDPJ regular / autos apartados	Benchmark humano	Acolhido no resultado processual	Determinante
Ausência de requisitos do art. 50	Benchmark humano + MVP + E2 + E4	Não decidido	Não determinante
Autonomia da sociedade limitada unipessoal	Benchmark humano + condições de IA defensivas	Não decidido	Não determinante

Fundamento	Quem apresentou	Tratamento judicial	Peso no resultado
Mera inexistência de bens insuficiente	Benchmark humano + condições de IA defensivas	Não decidido	Não determinante
Honorários	Benchmark humano + MVP + E4	Não decidido na decisão principal	Não determinante
Cronologia da constituição societária	MVP + E2	Não usada pelo juízo	Relevância defensiva ex ante, não decisória
Histórico de bem localizado	MVP + E2 + E4	Não usado pelo juízo	Relevância defensiva ex ante, não decisória

Controle de vies retrospectivo. O fundamento processual priorizado no benchmark humano coincidiu com o fundamento efetivamente usado pelo juízo, mas isso não transforma seus demais argumentos em padrão de correção. Da mesma forma, os fatos adicionais registrados pelo MVP não se tornam irrelevantes apenas porque o juízo decidiu em terreno processual. A comparação separa qualidade ex ante de proximidade ex post ao fundamento adotado.

12. Convergências, diferenças e acréscimos da condição MVP em relação ao benchmark humano

Dimensão	Conclusão
Convergências	Identificação do cliente e da fase; medida defensiva adequada; art. 50; autonomia societária; insuficiência de presunção automática; honorários; peça final coerente com a análise.
Menor ênfase ou aderência	Menor ênfase no núcleo processual posteriormente acolhido; exploração menos direta da distinção entre precedente de empresário individual e sociedade limitada; omissão do conflito de endereço; jurisprudência posterior ao corte histórico reduz comparabilidade.
Achados adicionais	Principal fato adverso cronológico; histórico do bem e inércia operacional do exequente; prova positiva de autonomia; distinção mais explícita entre prova e inferência; teste adversarial estruturado.

13. Falhas críticas por execução

- MVP / Execução 1: nenhuma falha fundacional identificada; correções localizadas. Os principais limites experimentais foram a assimetria temporal de jurisprudência e a omissão do conflito de endereço.
- Execução 2: falha crítica de integridade input-output ao converter citação incerta no corpus em fato certo na minuta.
- Execução 3: falha fundacional de identificação do cliente e erro de medida processual, produzindo peça contrária ao objetivo real.
- Execução 4: nenhuma falha fundacional, mas omissão material do principal fato adverso cronológico, além do conflito de endereço.
- Benchmark humano: sem falha de polo ou medida; fragilidades concentradas em prova, teste adversarial, precisão normativa e formulações categóricas sobre fatos não plenamente provados.

14. Limitações metodológicas

- Caso único. Não há base estatística para inferir desempenho geral, taxa de erro ou superioridade de uma abordagem.
- Caso selecionado retrospectivamente. A existência de decisão posterior e de um benchmark humano bem definido pode introduzir saliência específica que não existirá em outros litígios.
- Assimetria temporal de fontes jurídicas. Algumas IAs usaram precedentes posteriores ao momento histórico do benchmark humano. Esses pontos não são comparáveis como desempenho ex ante.
- Diferença entre corpus cego e exportação processual posterior. Fatos que ocorreram na realidade podem não ter estado disponíveis às execuções; o julgamento experimental deve seguir o input efetivamente fornecido.

- Benchmark humano não é padrão-ouro. A impugnação preexistente contém erros e omissões; resultado favorável não corrige essas falhas.
- Decisão judicial não é padrão-ouro jurídico completo. Ela revela o fundamento escolhido por um julgador, não necessariamente a melhor estratégia ou a correção de teses não enfrentadas.
- Dependência de revisão humana. Todas as condições defensivas exigiriam conferência de andamento atual e documentação interna antes de protocolo.
- O relatório cego agregou diagnósticos de execuções, mas não substitui logs completos, timestamps, parâmetros e versões de modelos se o objetivo futuro for reprodutibilidade técnica estrita.

15. Conclusões que os dados permitem

- Nesta rodada, a Execução 1 (MVP) apresentou a maior aderência global ao conjunto de critérios utilizado na auditoria cega.
- Nesta rodada, a condição MVP gerou análise e minuta processualmente adequadas após revisão, com aderência material em diversas dimensões comparadas.
- Nesta rodada, o MVP enfrentou explicitamente o principal fato adverso, recuperou fatos executivos relevantes, recomendou prova positiva e preservou de forma mais explícita a distinção entre prova e inferência.
- Nesta rodada, o benchmark humano se alinhou de forma mais direta ao fundamento processual que posteriormente decidiu o incidente e explorou com maior extensão a distinção entre empresário individual e sociedade limitada unipessoal.
- Outras execuções registraram sinais que não apareceram no benchmark humano nem no MVP, especialmente uma inconsistência entre documentos.
- Falhas críticas de IA podem ser fundacionais: assumir o cliente errado ou transformar incerteza em fato certo são erros de natureza diferente de mera redação ruim.

16. Conclusões que os dados ainda não permitem

- Não permitem afirmar que o MVP foi 'validado' para prática jurídica geral.
- Não permitem afirmar superioridade global do MVP sobre advogados ou sobre o benchmark humano em outros tipos de incidente.
- Não permitem estimar taxa de alucinação, taxa de omissão ou confiabilidade sem revisão humana.
- Não permitem inferir que o MVP produziria o mesmo resultado sob corpus diferente, sem jurisprudência posterior ou com outra parte representada.
- Não permitem afirmar que maior proximidade com a decisão posterior equivalha a maior qualidade jurídica ex ante.
- Não permitem separar com segurança o efeito do modelo, do prompt, do acesso a ferramentas e da organização do corpus, porque a rodada não isola estatisticamente esses componentes.

17. Próximos testes recomendáveis

1. Replicar a avaliação em conjunto maior e heterogêneo de casos reais, com diferentes ramos, fases e posições processuais.
2. Pré-registrar o corte temporal de fatos e de jurisprudência. Nenhuma execução deve usar precedente posterior à data histórica simulada.
3. Usar corpus idêntico, checksum/versão identificada e registro de quais documentos cada condição recebeu.
4. Manter avaliadores cegos quanto à identidade da condição e somente revelar o mapeamento após a pontuação independente.

5. Separar três produtos por execução: análise dos autos, plano estratégico e peça final. Avaliar integridade entre as três camadas.
6. Introduzir casos em que o melhor resultado é não peticionar, pedir esclarecimento ou produzir prova, para medir propensão a redigir prematuramente.
7. Incluir casos com contradições interdocumentais deliberadas, para testar cruzamento documental e não apenas leitura de documentos isolados.
8. Testar explicitamente identificação de cliente/polo com informação ausente, ambígua ou contraditória.
9. Medir preservação de incerteza do raciocínio para a peça final, com atenção a falsos fatos processuais.
10. Usar benchmark humano plural: mais de um advogado independente por caso, para evitar tratar uma única peça como padrão de excelência.
11. Avaliar separadamente qualidade ex ante e aderência ex post à decisão judicial, sem usar a segunda como substituto da primeira.
12. Relatar custo de correção humana por produto: nenhuma correção, correção localizada, correção material ou refeição integral.

18. Conclusão final da rodada

Os dados desta rodada mostram que a condição MVP produziu análise e minuta juridicamente utilizáveis após revisão, consideradas as exigências de controle de input, corte temporal, verificação de cliente e polo e conferência humana. O achado central não é uma vitória de uma condição sobre outra. O benchmark humano priorizou uma objeção processual coincidente com o fundamento decisório posterior; o MVP recuperou fatos materiais não explorados, estruturou o risco adverso e recomendou prova positiva. Uma terceira execução detectou inconsistência documental não registrada pelos outros dois produtos.

O experimento também registrou que os erros de IA não foram homogêneos: uma condição produziu peça para o polo errado, outra converteu incerteza em certeza, e condições com aderência em várias dimensões deixaram de registrar fatos materiais. Isso justifica ampliar a amostra antes de qualquer conclusão de validação. A rodada constitui evidência experimental inicial, auditável e informativa, mas insuficiente para alegações de desempenho geral.

Apêndice A - Pontos de reconstrução do experimento

- Preservar o arquivo dos autos fornecido às IAs e sua versão/corte temporal exatos.
- Preservar separadamente cada execução completa, inclusive mensagens de bloqueio/clarificação, análise e minuta final.
- Preservar o benchmark humano original e a decisão posterior em arquivos imutáveis.
- Preservar o relatório de auditoria cega e registrar quando a minuta da Execução 1 foi incorporada.
- Registrar explicitamente o mapeamento Execução 1 = MVP usado nesta fase final.
- Em futuras rodadas, registrar também versão do modelo, data de execução, ferramentas habilitadas e política de acesso à web/jurisprudência, sem alterar a queue do avaliador.

Apêndice B - Matriz de rastreabilidade das fontes

Ref.	Fonte	Conteúdo usado	Função
S1	Autos reservados	Pedido; cadastro; atos processuais; benchmark humano; resposta adversa e decisão posterior.	Documento processual primário.
S2	Auditoria cega consolidada	Método; condições; classificação interna; falhas e comparação transversal.	Relatório experimental consolidado.
S3	Auditoria do benchmark humano	Avaliação ex ante; decisão posterior; controle de viés; aderências e lacunas.	Produto analítico desta rodada.
S4	Comparação final	Síntese por dimensões com ajuste temporal e sem inferência geral.	Produto analítico.