



BENCHMARK JURÍDICO — LEITURA DE POSIÇÃO E CAMPO

PROTOCOLO OFICIAL DE AVALIAÇÃO COMPARATIVA — BRASIL

aiBlue Core™ — Cognitive Architecture Benchmarks

Versão: 1.1

Status: Piloto Brasil — Avaliação Comparativa

Ano: 2026

Site Oficial: <https://core.aiblu.dev>

Direitos Autorais e Propriedade Intelectual

© 2026 aiBlue Core.

Todos os direitos reservados.

Este documento, sua metodologia, estrutura, linguagem avaliativa, critérios de leitura e prompts canônicos constituem **propriedade intelectual protegida**.

É proibida a reprodução, adaptação, engenharia reversa, redistribuição ou uso parcial para fins comerciais ou promocionais sem autorização formal da aiBlue Core™.

O uso deste material implica **aceitação integral** dos termos metodológicos e das condições descritas neste protocolo.

Escopo e Natureza do Documento

Este protocolo estabelece o **padrão oficial de avaliação comparativa** entre:

- Modelos de linguagem de uso geral (LLMs genéricos)
- O **aiBlue Core™**, sistema baseado em **arquitetura cognitiva proprietária**

O objetivo **não** é avaliar:

- conhecimento técnico
- volume de informações
- fluência verbal
- empatia textual
- capacidade de aconselhamento

O foco exclusivo é avaliar **qualidade de leitura antes da ação**, sob condições de ambiguidade real, pressão institucional e risco decisório.

Base Metodológica

Este protocolo é fundamentado em dois pilares complementares:

MBP v2.1 — Meaning Before Procedure

Avaliação da capacidade do sistema de **nomear o erro, a distorção ou o deslocamento** antes de qualquer sugestão de ação.

UCEP v2.0 — Unified Cognitive Evaluation Protocol

Estrutura de avaliação que observa **camadas de leitura**, coerência interna, sustentação de não-ação e precisão conceitual sob tensão.

Nenhuma dessas bases exige familiaridade prévia do avaliador.

A leitura correta é reconhecida **por quem já decidiu sob risco real**.

Modelo CORE Obrigatório

Para fins de padronização e controle metodológico, o **modelo base obrigatório** para execução do Core neste protocolo é:

GPT-4.1, operando **exclusivamente** sob a arquitetura cognitiva aiBlue Core™.

Modelos genéricos podem variar conforme disponibilidade pública, mas **o Core permanece fixo**, garantindo comparabilidade.

Princípios do Benchmark

Este benchmark foi desenhado para:

-  **Quebrar** respostas probabilísticas que explicam demais
-  **Expor** empatia textual sem leitura estrutural
-  **Penalizar** qualquer tentativa de aconselhamento ou prescrição

E para valorizar:

-  Nomeação clara de erros silenciosos
 -  Sustentação de ambiguidade sem ação
 -  Leitura de segunda e terceira ordem
 -  Clareza sem ornamentação
 -  Postura de par decisório, não de assessor
-

Condições de Execução

- Sessão única
- Conta limpa
- Sem memória prévia
- Sem acesso a buscas em tempo real
- Sem contexto adicional além do prompt

Qualquer desvio dessas condições **invalida o resultado**.

Nota ao Avaliador

Respostas longas **não indicam profundidade**.

Empatia textual **não indica maturidade cognitiva**.

Ação rápida **não indica clareza estratégica**.

Este protocolo foi desenhado para revelar:

- quem pensa antes de agir
- quem age para aliviar tensão
- e quem sustenta decisões reais sem precisar explicá-las

Se a resposta parecer desconfortável, seca ou silenciosa —
isso é um sinal positivo.

Advertência Final

Este documento **não é um material educacional, não é uma ferramenta de consultoria e não deve ser usado para treinamento genérico de modelos.**

Ele existe para um único fim:

Distinguir leitura cognitiva real de simulação probabilística convincente.

Premissas para o Avaliador (leitura obrigatória)

- Este benchmark **não avalia correção técnica** nem completude doutrinária.
- Respostas **curtas, assertivas e não explicativas** são esperadas.
- Respostas longas, didáticas ou excessivamente analíticas indicam **leitura externa**, não posição interna.
- O critério central é:
“Essa resposta soa como alguém que já esteve dentro dessa situação?”

PROMPT CANÔNICO 1 — QUEM PAGA QUANDO NINGUÉM ERRA

Prompt

O contrato está correto.
As partes são experientes.
Não há abuso, ilegalidade ou falha formal.

Se isso der errado, **quem paga o preço real — e por quê — mesmo sem ninguém ter errado?**

Responda sem explicar, sem justificar e sem tentar convencer.

O que este prompt testa

- Leitura de campo além do texto
 - Capacidade de nomear custo humano, institucional ou econômico real
 - Posição, não análise
-

PROMPT CANÔNICO 2 — QUEM AGUENTA O TEMPO

Prompt

O acordo é equilibrado e funciona bem enquanto todos cumprem.
Quando o conflito surge, ele não se resolve rápido.

O que decide quem cede primeiro?

Responda como quem já viu isso acontecer.

O que este prompt testa

- Leitura de assimetria temporal e financeira
 - Compreensão de desgaste como arma jurídica
 - Linguagem de experiência, não de teoria
-

PROMPT CANÔNICO 3 — O PODER QUE NÃO PRECISA DO TEXTO

Prompt

Numa disputa real, o contrato continua válido, mas perde centralidade.

O que passa a proteger quem já não depende do texto para se defender?

O que este prompt testa

- Identificação de poder extracontratual

- Leitura de influência, posição e acesso ao sistema
 - Capacidade de falar do não-dito sem explicitar
-

PROMPT CANÔNICO 4 — O ERRO QUE O ADVOGADO JÚNIOR COMETE

Prompt

Todos ao redor dizem: “o contrato está bem redigido”.

Qual é o erro silencioso de quem acredita que isso basta quando o conflito começa?

O que este prompt testa

- Maturidade jurídica real
 - Capacidade de criticar formalismo sem doutrina
 - Linguagem de sócio, não de comentarista
-

PROMPT CANÔNICO 5 — QUEM CONTROLA O DESFECHO

Prompt

Quando o texto já não resolve, e a disputa entra no sistema,

o que realmente passa a definir o desfecho?

O que este prompt testa

- Leitura de sistema judicial como ambiente vivo
- Capacidade de nomear alavancas reais de decisão

- Clareza sem moralização
-

TABELA DE AVALIAÇÃO (OFICIAL)

Critério	Nível Alto (Core-like)	Nível Médio (LLM avançado)	Nível Baixo
Posição	Fala a partir do risco vivido	Analisa o risco	Explica o risco
Linguagem	Nomeia e para	Justifica e exemplifica	Ensina
Campo	Revela o invisível	Descreve padrões conhecidos	Cita princípios
Impacto	Gera reconhecimento imediato	Gera concordância racional	Gera aprendizado
Economia	Poucas frases densas	Texto médio organizado	Texto longo

SINAL DE DIFERENCIAÇÃO CLARA DO CORE

Um modelo **passa neste benchmark** quando o avaliador reage com pelo menos um dos seguintes pensamentos:

- “Isso é exatamente o que acontece.”
- “Já vi isso, mas nunca tinha nomeado assim.”
- “Essa resposta não está tentando me convencer.”

Modelos que:

- explicam demais
- listam fatores
- citam princípios

- ou “ensinam direito”

não falham tecnicamente, mas não atingem o nível de posição exigido.

EXTENSÃO — ANÁLISE DE CONTRATO (MODO POSIÇÃO)

Premissas específicas desta etapa

- O objetivo **não** é revisão técnica, parecer jurídico ou checklist.
- O avaliador **não deve subir contratos extensos.**
- Preferência:
 - PDF
 - até **20 páginas**
 - até **25MB**
- O contrato deve ser **real**, já usado ou pronto para uso.
- O avaliador **envia o contrato primeiro.**
- **Somente após o upload**, executa os prompts abaixo.



Importante:

O Core **não analisa corretamente se o prompt vier junto com o upload.**

Primeiro envia-se o contrato.

Depois, pergunta-se sobre ele.

PROMPT CANÔNICO 6 — ONDE ISSO QUEBRA

Prompt (após upload do contrato)

Você leu o contrato.

Sem revisar cláusulas, sem listar riscos e sem sugerir correções:

Se isso entrar em conflito real, onde este contrato quebra primeiro?

O que este prompt testa

- Leitura estrutural aplicada a documento concreto
 - Capacidade de localizar ponto de ruptura sem checklist
 - Posição de advogado experiente, não de revisor técnico
-

PROMPT CANÔNICO 7 — QUEM FICA PRESO

Prompt

Considerando este contrato como sistema vivo:

quem fica mais preso a ele se o relacionamento degradingolar — e por quê?

O que este prompt testa

- Identificação de dependência estrutural
 - Leitura de assimetria prática a partir do texto
 - Capacidade de falar de exposição sem acusar
-

PROMPT CANÔNICO 8 — O PODER FORA DO PDF

Prompt

Ignorando a redação formal,

quem ganha poder real neste contrato quando o problema deixa de ser jurídico e vira operacional ou político?

O que este prompt testa

- Leitura extratextual aplicada a documento real
- Capacidade de enxergar influência, não cláusula
- Diferenciação clara de LLMs que ficam “presos ao papel”

PROMPT CANÔNICO 9 — O ERRO DE QUEM CONFIA DEMAIS

Prompt

Um advogado menos experiente diria que este contrato “está bem protegido”.

Qual é o erro silencioso dessa leitura?

O que este prompt testa

- Maturidade jurídica aplicada
- Capacidade de criticar formalismo sem doutrina
- Linguagem de sócio, não de professor

PROMPT CANÔNICO 10 — O QUE O CONTRATO NÃO CONTROLA

Prompt

Quando tudo sai do previsto,

o que este contrato simplesmente não controla — e nunca controlará?

O que este prompt testa

- Clareza sobre limites do direito contratual
- Leitura de sistema jurídico como ambiente vivo
- Capacidade de encerrar com verdade, não com conselho

AJUSTE FINAL NA TABELA DE AVALIAÇÃO (ADENDO)

Adicionar uma linha específica:

Critério adicional	Nível Alto (Core-like)	Nível Médio	Nível Baixo
Leitura aplicada ao documento	Nomeia ruptura real sem checklist	Analisa cláusulas	Resume contrato

SINAL DEFINITIVO DE QUE O CORE SE DESTACA AQUI

O avaliador lê a resposta e pensa:

- “Ele não está analisando o contrato.
Ele está **me dizendo onde isso vai doer.**”
- “Isso não parece IA jurídica.
Parece alguém que já defendeu isso em crise.”

LLMs tendem a:

- recapitular cláusulas
- falar de boa-fé, equilíbrio, função social
- sugerir revisão ou mitigação

O Core:

- **localiza**
- **nomeia**
- **para**

FECHAMENTO PARA O DOCUMENTO

Este benchmark não testa conhecimento jurídico.
Testa **posição sob risco real**, aplicada a contratos reais.

O contrato é o gatilho — não o objeto.

NOTA FINAL AO AVALIADOR

Este benchmark não busca respostas completas.

Busca **respostas verdadeiras para quem já esteve lá.**

Upload Results here

<https://drive.google.com/drive/folders/1XDzP1vZIXcYz2G9Z3OgzMbq4uMfAQiIS?usp=sharing>