

MANUAL DE APLICAÇÃO DA MATRIZ ESCOLIA

Leonardo Fontes Vasconcelos¹

Orcid: 0009-0000-8202-3414

Instituto Lovelace de Inteligência Artificial Aplicada

Material originalmente publicado pelo autor na comunidade Filoia na Plataforma Zenodo ligada ao laboratório CERN da União Europeia.

Data da publicação original: 10 de março de 2026.

DOI da publicação original:

<https://doi.org/10.5281/zenodo.18942623>

Registro na Câmara Brasileira do Livro: ISBN 978-65-01-98730-9.

Técnica livre para uso público sem ônus.

Dê os devidos créditos ao autor.

Contato: contato@institutolovelace.org

¹ Diretor do Instituto Lovelace de Inteligência Artificial Aplicada. Professor universitário na Faculdade Anhanguera Rio Branco. Pós-Graduado em Direito Digital e especializado em inteligência artificial aplicada ao Direito. Membro vitalício da Academia de Letras Jurídicas do Estado do Acre. Membro Consultivo da Comissão de Inteligência Artificial da OAB Acre. Advogado licenciado e assessor no Ministério Público do Estado do Acre (MPAC). Coordenador de Pesquisa e Inovação do Centro de Estudos e Aperfeiçoamento Funcional do MPAC. Autor da primeira petição inicial elaborada com inteligência artificial protocolada no Judiciário. Coautor da Resolução 49/2023 da OAB, a primeira norma regulamentadora de IA na advocacia brasileira.

SUMÁRIO

APRESENTAÇÃO	3
COMO APLICAR A MATRIZ ESCOLIA	4
MÉTODO DE AVALIAÇÃO DOS CRITÉRIOS	5
Segurança de Dados	5
Oferta de Ferramentas Críticas	6
Estabilidade do Sistema	7
Precisão Analítica	7
Não-Alucinação	7
Janela de Contexto	8
Qualidade de Escrita.....	9
Outras Características Úteis	10
ÍNDICE GERAL.....	10
COMO ANALISAR E ESCOLHER	10
TABELAS DE CRITÉRIOS	10
CONSIDERAÇÕES FINAIS	12
ANEXO – EXEMPLO DE PREENCHIMENTO	13

MANUAL DE APLICAÇÃO DA MATRIZ ESCOLIA

Versão 1.3 | Março de 2026

APRESENTAÇÃO

A ascensão dos Grandes Modelos de Linguagem (LLMs) impulsionou a popularização da inteligência artificial em diversas esferas profissionais, impactando variados setores de forma distinta, mas estabelecendo-se como uma ferramenta transversal no mercado de trabalho.

O panorama nacional reflete essa tendência global: conforme os dados da pesquisa TIC Domicílios 2025, publicada pelo Comitê Gestor da Internet no Brasil (CGI.br) e pelo Cetic.br, o uso de ferramentas de IA generativa já é uma realidade para 32% dos usuários de internet no Brasil². Este dado evidencia que a tecnologia deixou de ser uma promessa futurista para se tornar um componente cotidiano da vida digital brasileira.

A experiência acumulada na condução de palestras e capacitações revela uma inquietação recorrente entre os profissionais: a dificuldade em identificar qual modelo de IA melhor atende às suas demandas específicas. É evidente que a escolha do modelo ideal é intrínseca à natureza da tarefa, bem como às expectativas e objetivos de desempenho de cada usuário em seu fluxo de trabalho.

Nesse contexto de incerteza técnica, a **Matriz Escolia**, desenvolvida pelo pesquisador e analista em inteligência artificial aplicada Leonardo Fontes Vasconcelos, surge como uma solução estruturada para usuários que demandam suporte e um método qualificado para a tomada de decisão. A ferramenta foi concebida para mitigar o caráter empírico e subjetivo da escolha tecnológica, substituindo a intuição por uma abordagem metodológica e sistematizada.

O objetivo central é conferir maior rigor metodológico ao processo de seleção, fundamentando-o em critérios técnicos e funcionais que sejam verdadeiramente relevantes para a atividade desenvolvida, assegurando uma aplicação mais estratégica e eficiente da inteligência artificial.

² Fonte: https://cetic.br/media/analises/tic_domicilios_2025_principais_resultados.pdf

O uso da **Matriz Escolia** privilegia um processo de escolha que seja justificável, auditável e transparente nas razões que levaram à escolha do modelo de IA selecionado.

Soma-se também o benefício de o registro do processo de análise e escolha pautado na Matriz possibilitar uma documentação de governança em inteligência artificial dos envolvidos.

COMO APLICAR A MATRIZ ESCOLIA

Tendo em vista que o objetivo é selecionar o modelo de LLM adequado para uma atividade em específico, e considerando que essa escolha se dá por meio de método mensurável, é fundamental que o analista (aquele que guiará o processo de seleção) institua um período de avaliação dos modelos em ambiente isolado e controlado que participarão da seleção, período comumente denominado *sandbox* e, se for para aplicação em uma instituição, é adequado que se forme um grupo de pessoas para avaliar os diversos modelos em análise.

Mesmo que não seja em uma corporação, que seja um usuário independente avaliando qual melhor IA aplicar em sua atividade, ainda assim o ideal é que se experimente diversos modelos para que se averigüe o de melhor desempenho para a tarefa pretendida.

Uma ressalva: é altamente sugerido que os testes avaliativos sejam feitos nos planos pagos de cada modelo de IA, pois eles possuem capacidades muito superiores em relação aos disponíveis no plano gratuito.

O critério de custo da assinatura não é objeto da Matriz. O avaliador deve previamente considerar o orçamento disponível para assinaturas de LLMs e então submeter à avaliação aqueles que já considera comportar em sua disposição financeira.

A **Matriz Escolia** apresenta alguns critérios avaliativos por padrão, mas o avaliador pode alterar ou acrescentar critérios de acordo com sua necessidade.

Os critérios padrões (e que já devem cobrir grande parte das atividades) são: segurança de dados, ferramentas críticas disponíveis, estabilidade do

sistema, precisão analítica, nível de alucinação (tratado como não-alucinação, como veremos adiante), janela de contexto e qualidade de escrita.

A análise resultará em resultados numéricos de cada critério e um **Índice Geral** do modelo na Matriz Escolia.

A Matriz independe de representação visual em formato de gráficos, contudo, construir **gráficos do tipo radar** facilita a visualização, comparação e tomada de decisão.

Veremos a seguir o método de análise de cada critério.

MÉTODO DE AVALIAÇÃO DOS CRITÉRIOS

Segurança de Dados

De maneira abstrata, esse é um dos critérios mais importantes, considerando que uma das maiores preocupações dos usuários de IA é quanto à segurança dos dados que se insere nos modelos.

A sua avaliação se dá na análise de itens que influenciam na visão global do nível de segurança de dados que o modelo de inteligência artificial oferece, são eles:

1. **Episódios de vazamento de dados:** realiza-se uma pesquisa por casos de vazamentos de dados ocorridos pelo modelo em análise. É interessante definir um recorte temporal para essa análise, não buscando desde o começo da existência do modelo, pois o nível de segurança de dados do modelo pode (e deve) evoluir com o tempo. Esse recorte temporal não é o mesmo do período de análise do modelo, deve ser superior como um ou dois anos. Guarde um registro dos episódios encontrados para fins de relatório posterior justificando a escolha tomada.
2. **Polêmicas de gestão:** também é realizada uma pesquisa para levantar os casos em que a gestão da empresa responsável pelo modelo analisado tenha tomado decisões controversas que podem (ao menos em tese) afetar o nível de segurança do modelo. Este é um item que guarda algum teor de subjetividade. Recomenda-se analisar o contexto da decisão de gestão para que se defina se de fato é relevante para a contagem e para o que se pretende do modelo analisado.

3. **Desativação de treinamento com os dados:** neste item, verifica-se se o modelo oferece a chave de controle para desativar o treinamento de novos modelos a partir dos dados dos chats do usuário.
4. **Existência de política expressa de segurança:** nem toda empresa possui política clara e transparente que prevê expressamente de maneira séria a forma com que gerencia a segurança de seus modelos.
5. **Existência de função de chat temporário:** este item verifica se o modelo oferece a função de chat temporário.

Cada item **deve ser avaliado de 0 a 1 ponto**. Os itens 3 a 5, estando presentes, levam 1 ponto. Diferente dos demais, os itens 1 e 2 possuem pontuação variável de 0 a 1, baseada na frequência de incidentes. A nota sofre uma redução progressiva conforme o volume de ocorrências: cada registro subtrai 0,1 do valor total. Dessa forma, caso sejam detectados três episódios de vazamento de dados, o item é pontuado com 0,7.

Ao final, deve-se fazer a soma da pontuação de cada item. A pontuação máxima nesse critério é de 5 pontos.

Oferta de Ferramentas Críticas

Este critério analisa se a plataforma oficial do modelo oferta ferramentas consideradas críticas e fundamentais para a tarefa pretendida.

São sugeridas como itens avaliativos, por padrão, as ferramentas **Projeto, Canvas (Lousa), Deep Research (Investigar), Instruções Personalizadas e Raciocinar (Pensar) por mais Tempo**.

O analista pode definir ou inserir outras ferramentas que considerar críticas para a atividade pretendidas.

A cada ferramenta que estiver presente, confere-se 1 ponto. A existência insuficiente (que não atende totalmente) gera uma pontuação dentro da escala entre 0 e 1.

Esse critério leva a pontuação na escala de 0 a 5 pontos, caso não haja inserção de novas ferramentas.

Estabilidade do Sistema

Esse critério analisa quantas falhas o sistema apresentou durante o período analisado. Podem ser casos de falhas generalizadas ou em funções específicas.

Observe que não se trata de erros de análise ou respostas incorretas, mas de falhas de funcionamento.

Comporta pontuação de 0 a 5 seguindo a progressão abaixo:

- Mais de 8 a 10 ocorrências: 1
- Mais de 6 a 8 ocorrências: 2
- Mais de 4 a 6 ocorrências: 3
- Mais de 2 a 4 ocorrências: 4
- De 0 até 2 ocorrências: 5

Precisão Analítica

A **Matriz Escolia** admite duas metodologias para a avaliação deste critério. A primeira consiste na condução de testes próprios pelo analista, visando mensurar o grau de assertividade do modelo em confronto com uma resposta de referência. Alternativamente, é possível adotar métricas de *benchmark* consolidadas, como o **AA-Omniscience** da Artificial Analysis, estabelecido como o padrão sugerido desta matriz.

A taxa de acurácia (*Accuracy*) do AA-Omniscience, equivalente à precisão analítica, vai de 0% a 100%.

A aplicação na Matriz deve se dar seguindo a seguinte progressão:

- Até 20%: 1
- Mais de 20% a 45%: 2
- Mais de 45% a 70%: 3
- Mais de 70% a 90%: 4
- Mais de 90% a 100%: 5

Não-Alucinação

Diferente do que faziam no início da popularização dos modelos de LLMs, a taxa de alucinação deixou de ser informada de maneira transparente

pelas desenvolvedoras dos modelos de IA. Ainda que elas continuassem informando, o ideal é usar métricas independentes para evitar manipulação e vieses institucionais nos dados.

Por essa razão, a Matriz também aplica o AA-Omniscience, utilizando a métrica *Hallucination Rate* reportada pelo *benchmark*. Como a *Hallucination Rate* mede a proporção de respostas incorretas dentre as tentativas de resposta do modelo, a Matriz usa a conversão apresentada pelo AA-Omniscience para esse valor em Não-Alucinação pela fórmula: $\text{Não-Alucinação} = 100\% - \text{Hallucination Rate}$. O usuário pode escolher outra métrica que achar mais adequada.

A taxa de Não-Alucinação, obtida pela conversão da *Hallucination Rate* do AA-Omniscience conforme descrito acima, vai de 0% a 100%.

A aplicação na Matriz deve se dar seguindo a seguinte progressão de acordo com o índice de Não-Alucinação:

- Até 20%: 1
- Mais de 20% a 45%: 2
- Mais de 45% a 70%: 3
- Mais de 70% a 90%: 4
- Mais de 90% a 100%: 5

Janela de Contexto

Cada modelo de IA possui uma janela de contexto com tamanho específico. Essa informação é apresentada pelo desenvolvedor do sistema.

O mais comum é que as janelas de contexto sejam entre 250 mil e 1 milhão de tokens.

A representação e registro deve se dar da seguinte maneira de acordo com o tamanho da janela de contexto anunciada:

- Até 300 mil: 1
- Mais de 300 mil a 500 mil: 2
- Mais de 500 mil a 700 mil: 3
- Mais de 700 mil a 1 milhão: 4
- Acima de 1 milhão: 5

Qualidade de Escrita

Este critério avalia a capacidade do modelo de produzir textos com qualidade redacional adequada à tarefa pretendida. A análise se fundamenta em dimensões específicas da produção textual que, em conjunto, oferecem uma visão objetiva sobre o nível de escrita do modelo.

As respostas geradas pelos modelos analisados devem ser avaliadas nos seguintes itens:

- **Coerência:** verifica se o texto apresenta progressão lógica de ideias, se os parágrafos se conectam adequadamente entre si e se a argumentação mantém consistência interna do início ao fim da resposta.
- **Fluência:** analisa se o texto é gramaticalmente correto, se utiliza vocabulário adequado ao registro esperado e se a leitura flui sem construções truncadas ou repetições desnecessárias.
- **Adequação ao contexto:** mensura se o modelo ajusta o nível de formalidade, a terminologia e o estilo ao tipo de documento ou resposta solicitada. Um parecer técnico exige registro diferente de uma comunicação informal, e o modelo precisa demonstrar essa sensibilidade.
- **Naturalidade de escrita:** avalia se o texto possui características de escrita natural ou se apresenta padrões artificiais típicos de geração automatizada, como repetição excessiva de estruturas frasais, uso formulaico de expressões de transição ou prolixidade sem conteúdo.
- **Estrutura e organização:** verifica se o texto apresenta uma arquitetura clara, com introdução, desenvolvimento e conclusão identificáveis, com paragrafação adequada e uso de conectivos que facilitem a compreensão.

Cada item deve ser avaliado de 0 a 1 ponto, considerando o desempenho médio do modelo nas respostas. A pontuação 1 indica atendimento pleno, 0 indica atendimento inexistente, e valores intermediários indicam atendimento parcial.

Outras Características Úteis

Opcionalmente, o analista pode utilizar este critério adicional para averiguar quantas outras características (ferramentas, funções e outros) ao uso proposto o modelo possui.

Caso seja utilizado esse critério, o fator de divisão do Índice Geral (apresentado abaixo) deve ser 8 no lugar de 7.

A escala sugerida é de 0 a 5, sendo 5 o limite independente de quantas características além de 5 o modelo possua.

ÍNDICE GERAL

Após ser encontrado o valor para cada um dos sete critérios, é calculado o Índice Geral do modelo analisado através da média da nota de todos os critérios. Soma-se a pontuação deles e divide-se por sete, seguindo a fórmula

$$\text{Índice Geral} = \text{soma das notas nos critérios} / 7$$

COMO ANALISAR E ESCOLHER

A Matriz Escolia comporta duas formas de escolha do modelo após a análise feita.

O analista pode escolher o modelo que auferiu a melhor média no **Índice Geral** ou o que obteve o melhor desempenho no conjunto de critérios mais importantes para a tarefa pretendida.

O anexo apresenta um exemplo da análise feita sobre o ChatGPT 5.4 Thinking, para que se compreenda a forma de preenchimento e análise.

TABELAS DE CRITÉRIOS

Para auxiliar o registro das informações durante a análise, foram construídas as tabelas a seguir para guiar a avaliação de cada modelo.

CRITÉRIO SEGURANÇA DE DADOS			
Modelo analisado			
Período de análise			
Item		Valor	Observação

Episódios de vazamento		
Polêmicas de gestão que fragilizam		
Desativar treinamento com os dados		
Política de segurança		
Chat temporário		
TOTAL DO CRITÉRIO		

CRITÉRIO FERRAMENTAS CRÍTICAS			
Modelo analisado			
Período de análise			
Ferramenta		Valor	Observação
Projetos			
Canvas (lousa)			
Deep Research			
Instruções Personalizadas			
Pensar por mais tempo			
TOTAL DO CRITÉRIO			

CRITÉRIO ESTABILIDADE DO SISTEMA				
Modelo analisado				
Período de análise				
1 (mais de 8 a 10 ou mais)	2 (mais de 6 a 8)	3 (mais de 4 a 6)	4 (mais de 2 a 4)	5 (até 2)
TOTAL DO CRITÉRIO				

CRITÉRIO PRECISÃO ANALÍTICA (Accuracy pela AA-Omniscience)				
Modelo analisado				
Período de análise				
1 (até 20%)	2 (mais de 20% a 45%)	3 (mais de 45% a 70%)	4 (mais de 70% a 90%)	5 (mais de 90% a 100%)
TOTAL DO CRITÉRIO				

CRITÉRIO NÃO-ALUCINAÇÃO (Não-Alucinação pela AA-Omniscience)				
Modelo analisado				
Período de análise				
1 (até 20%)	2 (mais de 20% a 45%)	3 (mais de 45% a 70%)	4 (mais de 70% a 90%)	5 (mais de 90% a 100%)
TOTAL DO CRITÉRIO				

CRITÉRIO JANELA DE CONTEXTO				
Modelo analisado				
Período de análise				
1 (até 300 mil)	2 (mais de 300 mil a 500 mil)	3 (mais de 500 mil a 700 mil)	4 (mais de 700 mil a 1 milhão)	5 (acima de 1 milhão)
TOTAL DO CRITÉRIO				

CRITÉRIO QUALIDADE DE ESCRITA		
Modelo analisado		
Período de análise		
Item		Valor de 0 a 1
Coerência		Observação
Fluência		
Adequação ao contexto		
Naturalidade de escrita		
Estrutura e organização		
TOTAL DO CRITÉRIO		

CONSIDERAÇÕES FINAIS

A **Matriz Escolia** constitui uma proposta metodológica destinada a transformar a escolha de sistemas de inteligência artificial generativa em um processo estruturado, transparente e justificável, substituindo decisões intuitivas por critérios comparáveis e auditáveis.

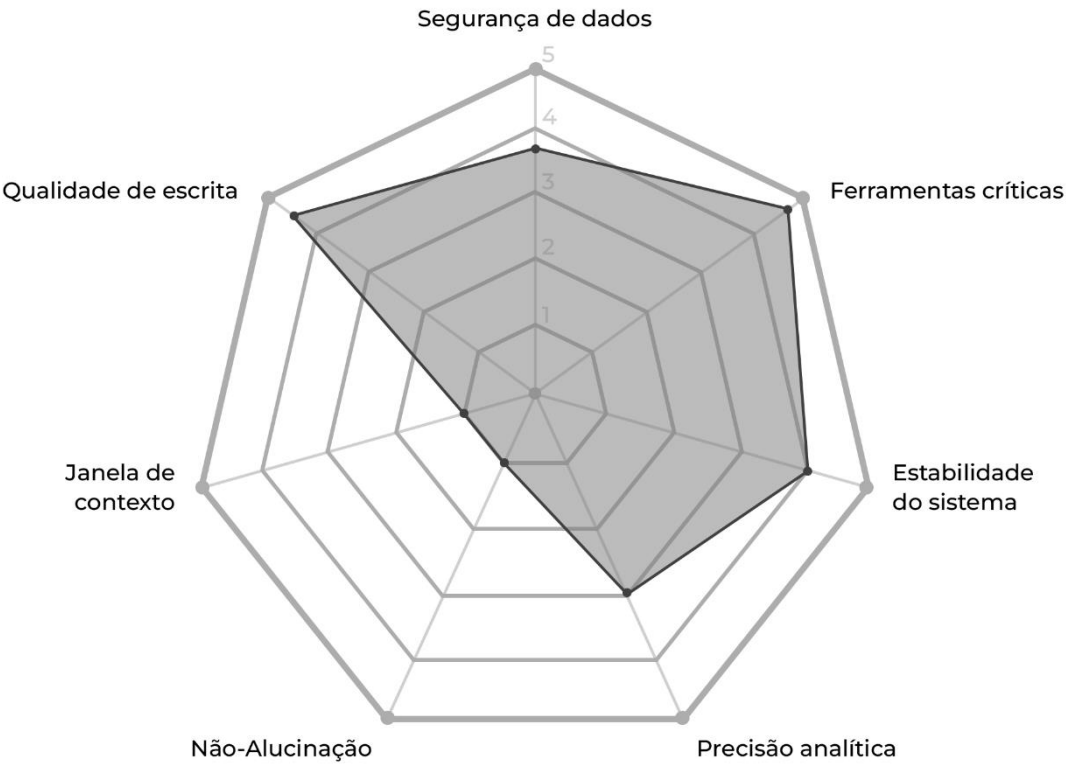
Ao integrar dimensões técnicas, como precisão analítica, estabilidade, taxa de alucinação, janela de contexto e qualidade de escrita, com aspectos de governança, como segurança de dados e disponibilidade de ferramentas críticas, o método oferece um instrumento prático para avaliar modelos de IA de acordo com as necessidades específicas de cada tarefa.

Nesse sentido, a Matriz não se limita a produzir rankings, mas estabelece um procedimento de análise que permite documentar, justificar e revisar decisões tecnológicas, contribuindo para uma adoção mais responsável, consciente e metodologicamente fundamentada da inteligência artificial em contextos profissionais e institucionais.



ANEXO – EXEMPLO DE PREENCHIMENTO

ChatGPT 5.4 Thinking



Índice Geral: 3,15

CRITÉRIO SEGURANÇA DE DADOS		
Modelo analisado	OpenAI ChatGPT 5.4 Thinking	
Período de análise	28/02/2026 a 09/03/2026	
Item	Valor	Observação
Episódios de vazamento	0,3	7 episódios
Polêmicas de gestão que fragilizam	0,5	5 polêmicas
Desativar treinamento com os dados	1	
Política de segurança	1	
Chat temporário	1	
TOTAL DO CRITÉRIO	3,8	

CRITÉRIO FERRAMENTAS CRÍTICAS		
Modelo analisado	OpenAI ChatGPT 5.4 Thinking	
Período de análise	28/02/2026 a 09/03/2026	
Ferramenta	Valor	Observação
Projetos	1	
Canvas (lousa)	0,8	Nem sempre responde em lousa
Deep Research	1	
Instruções Personalizadas	1	
Pensar por mais tempo	1	
TOTAL DO CRITÉRIO		4,8

CRITÉRIO ESTABILIDADE DO SISTEMA				
Modelo analisado	OpenAI ChatGPT 5.4 Thinking			
Período de análise	28/02/2026 a 09/03/2026			
1 (mais de 8 a 10 ou mais)	2 (mais de 6 a 8)	3 (mais de 4 a 6)	4 (mais de 2 a 4)	5 (até 2)
			x	
TOTAL DO CRITÉRIO				4,0

CRITÉRIO PRECISÃO ANALÍTICA (Accuracy pela AA-Omniscience)				
Modelo analisado	OpenAI ChatGPT 5.4 Thinking			
Período de análise	28/02/2026 a 09/03/2026			
1 (até 20%)	2 (mais de 20% a 45%)	3 (mais de 45% a 70%)	4 (mais de 70% a 90%)	5 (mais de 90% a 100%)
		x		
TOTAL DO CRITÉRIO				3,0

CRITÉRIO NÃO-ALUCINAÇÃO (Não-Alucinação pela AA-Omniscience)				
Modelo analisado	OpenAI ChatGPT 5.4 Thinking			
Período de análise	28/02/2026 a 09/03/2026			
1 (até 20%)	2 (mais de 20% a 45%)	3 (mais de 45% a 70%)	4 (mais de 70% a 90%)	5 (mais de 90% a 100%)
x				
TOTAL DO CRITÉRIO				1,0

CRITÉRIO JANELA DE CONTEXTO	
Modelo analisado	OpenAI ChatGPT 5.4 Thinking
Período de análise	28/02/2026 a 09/03/2026

1 (até 300 mil)	2 (mais de 300 mil a 500 mil)	3 (mais de 500 mil a 700 mil)	4 (mais de 700 mil a 1 milhão)	5 (acima de 1 milhão)
x				
TOTAL DO CRITÉRIO			1,0	

CRITÉRIO QUALIDADE DE ESCRITA		
Modelo analisado	OpenAI ChatGPT 5.4 Thinking	
Período de análise	28/02/2026 a 09/03/2026	
Item	Valor de 0 a 1	Observação
Coerência	1	
Fluência	1	
Adequação ao contexto	1	
Naturalidade de escrita	0,8	
Estrutura e organização	0,7	Os parágrafos tendem a ficar muito longos
TOTAL DO CRITÉRIO	4,5	