

TAREFA, COMPARAÇÃO E QUALIDADE

Checklist de avaliação de IA

Compare modelos na tarefa real da sua operação. Marque os critérios verificados e registre a decisão antes de colocar a solução em produção.

Não é um ranking nem um benchmark. Confira preços, limites e políticas na documentação oficial do provedor.

01 DEFINIR A TAREFA E O CRITÉRIO DE ACEITAÇÃO

- ☐ Descrever a tarefa e o resultado esperado antes de escolher o modelo.
- ☐ Separar exemplos de teste dos exemplos usados para ajustar o prompt.
- ☐ Incluir entradas típicas, ambíguas, incompletas e tentativas de instrução maliciosa.
- ☐ Definir quais erros impedem a adoção e quando uma pessoa deve assumir.

02 IDENTIFICAR EXATAMENTE O QUE ESTÁ SENDO COMPARADO

- ☐ Registrar provedor, ID/versão do modelo, data, região e parâmetros.
- ☐ Manter o mesmo conjunto de entradas, ferramentas e critérios de avaliação.
- ☐ Registrar mudanças de prompt, versão ou infraestrutura entre as execuções.
- ☐ Conferir documentação oficial sobre limites de contexto e disponibilidade.

03 MEDIR QUALIDADE E DESEMPENHO

- ☐ Avaliar respostas corretas, afirmações sem fonte e necessidade de retrabalho.
- ☐ Verificar estrutura de saída e argumentos de ferramentas antes da execução.
- ☐ Medir tempo até a primeira resposta e tempo total, incluindo ferramentas.
- ☐ Comparar mediana e percentil 95 sob a concorrência esperada, não só a melhor chamada.
- ☐ Repetir os casos para observar variação e registrar falhas e limites de uso.

Custo, segurança e operação

Não é um ranking nem um benchmark. Confira preços, limites e políticas na documentação oficial do provedor.

04 CALCULAR CUSTO POR TAREFA CONCLUÍDA

- ☐ Consultar o preço oficial vigente de entrada, saída e cache, quando aplicável.
- ☐ Incluir chamadas de ferramentas, busca, infraestrutura e revisão humana.
- ☐ Contabilizar novas tentativas e execuções que falharam.
- ☐ Dividir o custo total pelo número de tarefas aceitas pelo critério definido.
- ☐ Diferenciar assinatura, franquia de uso e cobrança por API; nenhum é capacidade ilimitada.

05 PRIVACIDADE E SEGURANÇA

- ☐ Usar dados sintéticos ou anonimizados nos testes iniciais.
- ☐ Conferir retenção, treinamento com dados, localização e contrato do provedor.
- ☐ Definir permissões mínimas; exigir aprovação para ações financeiras ou irreversíveis.
- ☐ Não enviar senhas, tokens nem dados de clientes sem base e controles adequados.
- ☐ Testar vazamento de dados entre clientes e instruções maliciosas em documentos.

06 DECIDIR E OPERAR

- ☐ Documentar os resultados observados e os casos em que o modelo não serve.
- ☐ Comparar o ganho com a alternativa atual, incluindo uma solução sem IA.
- ☐ Definir limites de custo, responsável por incidentes, revisão e reversão.
- ☐ Reavaliar ao mudar o modelo, prompt, dados ou conjunto de ferramentas.

CHECKLIST DE AVALIAÇÃO DE IA

Ficha de avaliação

Preencha uma ficha por modelo e versão. Use o mesmo conjunto de teste para comparar os resultados.

Tarefa e responsável:

Data / provedor / ID do modelo:

Conjunto de teste e critérios:

Tarefas aceitas / tarefas avaliadas:

Tempo mediano / percentil 95:

Custo total / tarefas aceitas:

Riscos, restrições e decisão:

Quer avaliar a aplicação no seu negócio?

Conheça a consultoria em IA →

openmidia.agency/servicos/consultoria-ia

Material gratuito. Sem cadastro ou obrigação de contratar.