



Orientação de correção. Os critérios abaixo descrevem respostas esperadas e evidências de domínio conceitual, não uma redação obrigatória. Aceite formulações alternativas tecnicamente corretas e bem fundamentadas. Distribua crédito parcial pelos argumentos corretos, mesmo quando a conclusão final diferir.

Questão 1 — Triagem com dados heterogêneos

(5,0 pontos)

- **Item (a) — Diagnóstico da Figura 1 e Decisão de Produção (2,0):** interpreta que a perda de treino caindo monotonicamente até a época 30 enquanto a perda de validação descola e sobe a partir da Época 8 diagnostica sobreajuste (*overfitting*/memorização) (1,0); refuta a premissa da equipe, apontando que queda na perda de treino isolada não garante generalização e causará falhas críticas em novos pacientes (1,0).
- **Item (b) — Representação e Retropropagação (2,0):** normaliza/escalone variáveis contínuas (idade, sinais vitais) e codifica categorias adequadamente (one-hot ou embedding), preservando rigorosamente as transformações na inferência (1,0); fundamenta que a retropropagação calcula os gradientes da perda em relação a pesos e vieses via regra da cadeia, e o otimizador atualiza os parâmetros (a rede não “aprende sozinha”) (1,0).
- **Item (c) — Evidências de Confiança e Reprodutibilidade (1,0):** exige validação/teste independentes sem vazamento, métricas clínicas adequadas (sensibilidade/recall, curva PR, calibração) e medidas concretas de reprodutibilidade (fixação de sementes e versionamento de pipeline) (1,0).

Alinhamento Didático com o Livro-Texto e Laboratórios:

- **Livro Didático:** Cap. 02 (*MLP*), Cap. 03 (*Descida do Gradiente*) e Cap. 05 (*Diagnóstico e Overfitting*).
- **Laboratórios:** Lab 02 (*MLP manual*), Lab 03 (*Convergência e Perda*) e Lab 05 (*Diagnósticos de Sobreajuste*).
- **Questão Reflexiva no Livro: Capítulo 05, Questão 4:** “Uma equipe técnica treinou um classificador de diagnóstico médico com 99% de acurácia no treino. A gerência quer colocá-lo imediatamente em operação hospitalar. Quais argumentos técnicos e metodológicos você apresentaria para exigir validação prévia em dados de teste independentes?”

Questão 2 — Inspeção visual em condições novas

(5,0 pontos)

- **Item (a) — Diagnóstico Experimental (2,0):** propõe coletar e inspecionar amostra rotulada do turno noturno analisando iluminação/ruído antes de alterar a rede (1,0); distingue com clareza *sobreajuste* (alto treino, baixa validação em dados similares), *mudança de distribuição* (*covariate shift* restrito à noite) e *falha de rotulação* (discordância sistemática com o especialista) (1,0).
- **Item (b) — Arquitetura e Transferência de Aprendizado (2,0):** fundamenta a CNN pela invariância a translação e extração de padrões locais (trincas/defeitos) (1,0); desenha estratégia de *transfer learning* congelando extratores pré-treinados, treinando nova cabeça e aplicando regularização e *data augmentation* realista (brilho/ruído) (1,0).
- **Item (c) — Cautela Amostral (1,0):** exige validação independente para o turno noturno, intervalos de confiança ou validação cruzada/bootstrap, evitando declarar ganho com base em conjunto diminuto do novo turno (1,0).

Alinhamento Didático com o Livro-Texto e Laboratórios:

- **Livro Didático:** Cap. 06 (*Regularização, Data Augmentation e Callbacks*) e Cap. 07 (*Transfer Learning e Otimizadores*).
- **Laboratórios:** Lab 07 (Partes 1 e 2 — *Classificador MobileNetV2, Congelamento de Camadas e Telemetria MLOps*).
- **Questão Reflexiva no Livro: Capítulo 07, Questão 4:** “Explique o fenômeno de *Shortcut Learning* em redes neurais de visão computacional. Como um engenheiro de IA pode desenhar um teste experimental (*stress test*) para comprovar se seu modelo realmente aprendeu a morfologia dos objetos ou se está apenas explorando viés de contexto?”

Questão 3 — Alerta que muda decisões

(5,0 pontos)

- **Item (a) — Leitura da Calibração e Crítica ao Limiar (2,0):** aponta que 0,50 é convenção simétrica arbitrária; na Figura 2, para $\hat{p} = 0,50$, a frequência real de urgência é de apenas 18%, provando sobreconfiança e disparo massivo de alarmes falsos (1,0); explica que calibrar exige que $\hat{p}$ reflita a frequência empírica observada ($y = x$), propondo calibração pós-processamento (Platt Scaling ou Regressão Isotônica) (1,0).
- **Item (b) — Custos Assimétricos e Capacidade Operacional (2,0):** formaliza que o custo de Falsos Negativos (óbito/urgência desassistida) supera o de Falsos Positivos (1,0); pondera a escolha do limiar pelo compromisso entre sensibilidade e o teto de vazão da equipe médica para evitar fadiga de alarmes (1,0).
- **Item (c) — Telemetria e Deriva de Conceito (1,0):** projeta monitoramento com métricas de decisão diárias, rastreamento de *data drift* e *concept drift*, com ações de re-calibração ou revisão médica diante de anomalias (1,0).

Alinhamento Didático com o Livro-Texto e Laboratórios:

- **Livro Didático:** Cap. 08 (*Avaliação Crítica, Métricas de Classificação, Matriz de Confusão e MLOps*).
- **Laboratórios:** Lab 08 (Partes 1 e 2 — *Avaliação Crítica, Serialização de Modelos e API de Produção com FastAPI/Docker*).
- **Questão Reflexiva no Livro: Capítulo 08, Questão 4:** “Em um sistema de detecção de fraudes [ou triagem crítica], é pior um Falso Positivo (bloquear uma transação legítima / alarme falso que sobrecarrega a equipe) ou um Falso Negativo (deixar uma emergência escapar)? A resposta muda se o sistema for de diagnóstico de saúde/UTI?”

Questão 4 — Busca de conhecimento para atendimento

(5,0 pontos)

- **Item (a) — Representação Semântica e Limitações do Cosseno (2,0):** distingue embeddings estáticos (Word2Vec: vetor fixo por palavra, ignora contexto) de contextualizados (Transformers/BERT: vetores variam conforme sintaxe e semântica da consulta) (1,0); demonstra que o cosseno reflete apenas proximidade angular, sem garantir autoridade, recência, veracidade factual ou termos exatos (1,0).
- **Item (b) — Avaliação Rigorosa de Recuperação (1,5):** rejeita a métrica ingênua (“retornou algo”); propõe conjunto de consultas anotadas com *ground truth* e métricas formais de IR como Recall@K, Precision@K, MRR ou NDCG (1,5).
- **Item (c) — Busca Híbrida e Governança (1,5):** combina busca vetorial densa com busca lexical (BM25) e filtros de metadados estruturados (vigência/data), assegurando que normas revogadas não superem normas vigentes (1,5).

Alinhamento Didático com o Livro-Texto e Laboratórios:

- **Livro Didático:** Cap. 09 (*Espaço Latente e Word Embeddings*) e Cap. 10 (*Busca Semântica, Similaridade de Cosseno e Busca Híbrida*).
- **Laboratórios:** Lab 09 (*Embeddings em Português*) e Lab 10 (Partes 1 e 2 — *Buscador Semântico, Ground Truth e Métricas de IR*).
- **Questão Reflexiva no Livro: Capítulo 10, Questão 4:** “A busca semântica ‘resolve’ sinonímia, mas pode falhar em domínios altamente especializados (ex: jargão técnico/normativo ou termos exatos). Proponha uma estratégia para mitigar esse problema usando conceitos deste capítulo.”

Questão 5 — Assistente que pode não responder

(5,0 pontos)

- **Item (a) — Contestação do Maior Escore e Desenho da Abstenção (2,0):** contesta a proposta simplista: em perguntas fora de escopo ou ambíguas, o maior escore pode ter baixa confiança e induzir alucinação (1,0); projeta política formal com duplo limiar: emite resposta se escore máximo $s_1 \geq \tau_1$ E margem $\Delta = s_1 - s_2 \geq \tau_2$, abstendo-se ou encaminhando a humano no caso contrário (1,0).
- **Item (b) — Conflito de Fontes e Validação Prévia (2,0):** estabelece regras de proveniência e hierarquia temporal para normas conflitantes (1,0); valida experimentalmente com consultas respondíveis e intencionalmente não respondíveis/ambíguas, mapeando cobertura vs. taxa de erro (1,0).
- **Item (c) — Telemetria e Impacto Operacional (1,0):** monitora a taxa contínua de abstenção em produção; aumentos na taxa alertam sobre mudanças nas dúvidas dos usuários (*query drift*) ou lacunas do acervo após alterações normativas, direcionando curadoria (1,0).

Alinhamento Didático com o Livro-Texto e Laboratórios:

- **Livro Didático:** Cap. 10 (Parte 2 — *Engenharia de Recuperação, Abstenção e Proveniência*).
- **Laboratórios:** Lab 10 (Parte 2 — *Abstenção com Thresholding, Ground Truth e Retorno Auditável*).
- **Questão de Revisão no Livro: Capítulo 10, Questão 3:** “Explique por que não se deve decidir um limiar de relevância apenas porque um documento recebeu cosseno 0.92. Que dados e métricas você usaria para calibrá-lo?” (conforme fundamentado na seção de RAG Responsável — *Abstenção e Citações* do Cap. 10).