Laudo Oficial de Correção — Prova Teórica 1

Inteligência Artificial II (IA II) ▪ CEFET-MG 2026.2

Prof. Aléssio Miranda Júnior

2026-10-19

Dados do Discente

- **Discente:** Daniel Caetano de Assis
- **Matrícula:** 20233003990
- **Turma:** Turma 2
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 24,9 / 25,0** (Aproveitamento: **99,6%**)
- **Legibilidade da Caligrafia:** 95% (Alta Legibilidade)

Desempenho Consolidado por Questão

Questão / Tema	Pontos Obtidos	Valor	Aproveitamento
Q1: Triagem com dados heterogêneos	**5,0**	5,0	100,0%
Q2: Inspeção visual em condições novas	**4,9**	5,0	98,0%
Q3: Alerta que muda decisões	**5,0**	5,0	100,0%
Q4: Busca de conhecimento para atendimento	**5,0**	5,0	100,0%
Q5: Assistente que pode não responder	**5,0**	5,0	100,0%
TOTAL GERAL	**24,9**	**25,0**	**99,6%**

Questão 1 — Triagem com dados heterogêneos (5,0 / 5,0 pts)

1.a Diagnóstico da Curva de Aprendizado e Decisão de Produção (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Afirmação falsa. Após a época 8 a perda de treino cai mas a de validação sobe: overfitting, o modelo decora o treino e não generaliza. Perda de treino baixa não garante desempenho em dados novos. O ideal seria parar na época 8 e avaliar no teste."

Parecer do Docente

Resposta impecável. Identificou com precisão a bifurcação na Época 8, explicou o mecanismo do sobreajuste (memorização versus generalização), vetou a ida a produção e sugeriu corretamente a política de early stopping.

1.b Representação de Entradas Heterogêneas e Retropropagação (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Numéricas: Padronizar (z-score) com estatísticas do treino. Categóricas: One-hot. Backpropagation calcula o gradiente de perda em relação a cada peso. O otimizador usa esses gradientes para atualizar os pesos."

Parecer do Docente

Rigor exemplar! Notou o detalhe crítico de engenharia: calcular o Z-score com estatísticas estritamente restritas ao conjunto de treino (evitando vazamento de dados), aplicou One-Hot Encoding e discriminou com clareza o papel do backpropagation (cálculo de gradientes via regra da cadeia) e do otimizador.

1.c Evidências Clínicas de Confiança e Reprodutibilidade (1,0 / 1,0)

Resposta Transcrita do Aluno

> "Divisão treino/validação/teste estratificada e sem vazamento, teste usado uma vez. Métricas: recall dos urgentes, precisão, matriz de confusão, AUC-PR e calibração. Reprodutibilidade: fixar a semente e versionar dados e splits."

Parecer do Docente

Padrão de produção completo: partição estratificada com teste blind intocado, métricas assimétricas vitais (Recall de urgentes, AUC-PR e calibração) e reprodutibilidade via fixação de sementes e versionamento de splits de dados.

Questão 2 — Inspeção visual em condições novas (4,9 / 5,0 pts)

2.a Diagnóstico Experimental do Aumento de Alertas (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Revisar amostra dos alertas noturnos e comparar métricas e distribuição das imagens dia x noite. Overfitting: erra também em imagens novas do dia. Covariate shift: vai bem de dia, erra à noite e as imagens noturnas têm outra distribuição. Rotulação: a revisão acha peças aprovadas com defeito ou discordância entre inspectadores."

Parecer do Docente

Excelente formulação experimental! Propôs o teste de contraste diurno vs noturno e delineou a assinatura empírica inequívoca de cada uma das três hipóteses.

2.b Arquitetura Convolucional e Transferência de Aprendizado (1,9 / 2,0)

Resposta Transcrita do Aluno

> "CNN capta padrões locais (bordas, texturas), compartilha pesos e é invariante à translação. Congelar a base pré-treinada e treinar só a cabeça nova. Depois ajuste fino das últimas camadas com learning rate baixo. Usar dropout, L2, early stopping e augmentation de brilho, constraste, ruído e rotação."

Parecer do Docente

Fundamentação teórica sólida (compartilhamento de pesos e invariância translacional) aliada a um plano prático completo de transfer learning, regularização (dropout e L2) e aumento de dados fotométrico. Levíssima dedução pela ausência de esquema gráfico.

2.c Cautela Amostral e Validação com Poucos Dados (1,0 / 1,0)

Resposta Transcrita do Aluno

> "Validação cruzada estratificada, média ± intervalo de confiança, teste noturno separado e intocado, comparar com o modelo atual. Ganho pequeno pode ser ruído."

Parecer do Docente

Excelente visão estatística: alertou expressamente sobre o risco de ganhos espúrios fruto de ruído amostral e instituiu intervalos de confiança com partição noturna cega.

Questão 3 — Alerta que muda decisões (5,0 / 5,0 pts)

3.a Interpretação da Calibração e Crítica ao Limiar 0,50 (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Em $\hat{p} = 0,5$ a frequência real é só 18%: o modelo superestima o risco, então em 0,5 ele erra $\approx 82\%$. Mal calibrado = a probabilidade não corresponde à frequência real e não pode ser lida como risco, é preciso recalibrar."

Parecer do Docente

Leitura quantitativa perfeita da Figura 2: calculou o erro empírico de 82% no limiar 0,5, diagnosticou a superestimação de risco e explicitou a impossibilidade de usar probabilidades nominais não calibradas na tomada de decisão médica.

3.b Custos Assimétricos e Capacidade da Equipe Médica (2,0 / 2,0)

Resposta Transcrita do Aluno

> "FN custa mais que FP, o que puxa o limiar para baixo, mas FP em excesso estoura a capacidade de equipe e gera fadiga de alarme. Escolher na validação o limiar que minimiza $C_{FN} \cdot FN + C_{FP} \cdot FP$ com alarmes $\leq$ capacidade."

Parecer do Docente

Formalização de alto nível! Expressou o problema como otimização de custo esperado sob restrição de capacidade física da equipe médica, prevenindo o fenômeno de fadiga de alarmes.

3.c Telemetria Contínua e Deriva de Conceito (1,0 / 1,0)

Resposta Transcrita do Aluno

> "Monitorar a distribuição das entradas (PSI/KS), dos escores, taxa de alarmes e, com desfechos, recall e calibração. Ao detectar anomalia: revisão humana, voltar a versão/limiar seguro, recalibrar ou retreinar com dados novos."

Parecer do Docente

Uso exemplar de métricas industriais de monitoramento de deriva (PSI e teste de Kolmogorov-Smirnov KS), associadas a um plano de governança com rollback seguro e retreino.

Questão 4 — Busca de conhecimento para atendimento (5,0 / 5,0 pts)

4.a Representação Semântica e Limitações da Similaridade de Cosseno (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Word2Vec: vetor fixo por palavra, ignora contexto. BERT: vetor depende da frase e capta paráfrases, mas custa mais e tem limites de tokens. Cosseno mede proximidade do tema, não relevância: sempre retorna algo, ignora se a norma é vigente, falha em códigos e termos exatos."

Parecer do Docente

Contraste conceitual impecável entre embeddings estáticos e representações contextuais profundas. Identificou com precisão cirúrgica os três pontos cegos da similaridade de cosseno (forçar retorno, cegueira à vigência e ineficácia em buscas exatas/códigos).

4.b Avaliação Rigorosa de Recuperação (Information Retrieval) (1,5 / 1,5)

Resposta Transcrita do Aluno

> "Busca vetorial sempre retorna resultado, então a taxa fica $\approx 100\%$ mesmo errando. Protocolo: consultas com ground truth, validação e testes separados. Precision@K, Recall@K, MRR e nDCG, comparando com BM25."

Parecer do Docente

Desarmou com perfeição a ilusão da métrica ingênua de "taxa de retorno" e estabeleceu a bateria completa de avaliação de IR (MRR, nDCG, Recall@K e Precision@K com baseline BM25).

4.c Busca Híbrida e Governança Normativa (1,5 / 1,5)

Resposta Transcrita do Aluno

> "BM25 acerta termos exatos e códigos, o denso acerta paráfrases; combinar os escores. Filtrar por metadados (data, versão, ...) antes do ranking, excluindo normas revogadas e citando fonte e versão."

Parecer do Docente

Arquitetura de busca híbrida descrita com precisão profissional, combinando fusão de escores com pré-filtragem obrigatória por metadados para garantir conformidade legal e citação auditável.

Questão 5 — Assistente que pode não responder (5,0 / 5,0 pts)

5.a Contestação do Maior Escore e Desenho da Abstenção (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Contesto: o maior score sempre devolve algo, mesmo sem resposta real. Política: se $s_1 < \tau_1$, abster; se $s_1 - s_2 < \tau_2$ e os documentos divergem, abster ou pedir esclarecimento; senão responder citando a fonte. τ_1 e τ_2 ajustados na validação."

Parecer do Docente

Formulação matemática e lógica de decisão irrepreensíveis! Articulou o limiar de pertinência mínima e a margem de segurança contra ambiguidade , prevendo desambiguação e citação transparente de fontes.

5.b Conflito de Fontes e Validação Prévia da Abstenção (2,0 / 2,0)

Resposta Transcrita do Aluno

> "Conflito: priorizar a norma vigente e mais recente via metadados; se persistir, abster e sinalizar. Validar com consultas respondíveis e não respondíveis, medindo acerto, abstenção correta e abstenção indevida para escolher τ_1 e τ_2."

Parecer do Docente

Excelente rigor de governança: priorização por vigência e recência, com matriz de confusão especializada para sistemas com abstenção (abstenção correta versus indevida) balizando a calibração dos limiares.

5.c Telemetria da Taxa de Abstenção em Produção (1,0 / 1,0)

Resposta Transcrita do Aluno

> "Registrar escores e decisão por consulta e acompanhar a taxa de abstenção. Se sobe: perguntas novas ou acervo desatualizado (reindexar). Se cai bruscamente: possíveis respostas erradas (auditar)."

Parecer do Docente

Análise bidirecional primorosa: associou o aumento na taxa de recusa à defasagem da base de conhecimento e interpretou com agudeza a queda abrupta da taxa como sintoma perigoso de alucinação ou relaxamento indevido de segurança.
