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:** Arthur Baltar Eler
- **Matrícula:** 20223004240
- **Turma:** Turma 2
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 16,5 / 25,0** (Aproveitamento: **66,0%**)
- **Legibilidade da Caligrafia:** 88% (Legível)

Desempenho Consolidado por Questão

Questão / Tema	Pontos Obtidos	Valor	Aproveitamento
Q1: Triagem com dados heterogêneos	**3,9**	5,0	78,0%
Q2: Inspeção visual em condições novas	**3,1**	5,0	62,0%
Q3: Alerta que muda decisões	**2,4**	5,0	48,0%
Q4: Busca de conhecimento para atendimento	**3,4**	5,0	68,0%
Q5: Assistente que pode não responder	**3,7**	5,0	74,0%
TOTAL GERAL	**16,5**	**25,0**	**66,0%**

Questão 1 — Triagem com dados heterogêneos (3,9 / 5,0 pts)

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

Resposta Transcrita do Aluno

> "A afirmação da equipe está equivocada, como nota-se no gráfico, o modelo aceita ~99,9% nessa base, porém na validação, a partir da época 8 revela que o modelo praticamente 'decorou' as respostas e não consegue generalizar, por isso a queda contínua não garante que o modelo esteja pronto para produção."

Parecer do Docente

Excelente resposta. O discente identificou com clareza o ponto de inflexão na Época 8 (onde a perda de validação volta a crescer enquanto a de treino declina), caracterizando com precisão o sobreajuste (overfitting / memorização) e refutando categoricamente a afirmação ingênua da equipe.

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

Resposta Transcrita do Aluno

> "É desejável que as entradas sejam normalizadas para alimentar corretamente a MLP. O backpropagation por sua vez serve para ajustar os pesos durante o treinamento, pois manter fixo diminui os ganhos, e a generalização do modelo."

Parecer do Docente

Correto ao prescrever normalização para atributos numéricos e ao situar a retropropagação no ajuste de pesos. Contudo, a resposta omitiu o tratamento obrigatório de variáveis categóricas (one-hot encoding ou embeddings) e não detalhou o mecanismo matemático da retropropagação (cálculo de gradientes via regra da cadeia a partir da função de perda para atualização pelo otimizador).

1.c Evidências Clínicas de Confiança e Reprodutibilidade (0,7 / 1,0)

Resposta Transcrita do Aluno

> "Uma medida concreta para tal é o recall, pois priorizará os casos mais graves, dando prioridade de atendimento para esses. Com respeito às evidências adicionais, particionar os dados corrobora a melhoria, pois previne/mitiga o problema de 'decorar respostas'. Já com relação às métricas de clínica, são úteis pois servem para comparar e avaliar a saída do modelo."

Parecer do Docente

Pertinente ao selecionar a métrica de recall voltada à minimização de falsos negativos em triagem e ao apontar o particionamento de dados. Faltou explicitar garantias concretas de reprodutibilidade experimental (fixação de sementes aleatórias e versionamento de artefatos de pré-processamento/pipeline com MLflow).

Questão 2 — Inspeção visual em condições novas (3,1 / 5,0 pts)

2.a Diagnóstico Experimental do Aumento de Alertas (1,2 / 2,0)

Resposta Transcrita do Aluno

> "Primeiro investigar as possibilidades: overfitting, covariate shift e outros. Se congelarmos as primeiras camadas e o erro persistir, significa que não é overfitting, se trocarmos a base verificamos a mudança de distribuição e tomamos conclusões com base na saída."

Parecer do Docente

Reconheceu nominalmente as hipóteses de sobreajuste e mudança de distribuição (covariate shift). Todavia, o plano experimental sugerido (congelar camadas para diagnosticar) é conceitualmente inadequado para inspeção de causa raiz; o procedimento exigia coletar e inspecionar amostras rotuladas do turno noturno (avaliando iluminação, ruído e alinhamento com os inspetores humanos) antes de qualquer manipulação de modelo.

2.b Arquitetura Convolucional e Transferência de Aprendizado (1,5 / 2,0)

Resposta Transcrita do Aluno

> "Uma CNN é apropriada pois treinar do zero uma rede neural pra imagens é custoso computacionalmente. Usando uma base pré-treinada (transfer learning) ajuda. Porém, seria interessante congelar as primeiras camadas logo no início pra não ocorrer overfitting e deixar a head livre, que seria o ajuste fino pra identificar as peças produzidas. Por fim, regularizar a base ajudaria, pois popularizaria a base que pode estar viciando o modelo."

Parecer do Docente

Muito boa formulação sobre a estratégia prática de transfer learning (congelamento das camadas extratoras pré-treinadas e fine-tuning da cabeça densa). A fundamentação da CNN, entretanto, limitou-se ao argumento de custo computacional, deixando de explicar a propriedade técnica primordial: a invariância por translação e os campos receptivos locais que aprendem hierarquias visuais (bordas, texturas, trincas).

2.c Cautela Amostral e Validação com Poucos Dados (0,4 / 1,0)

Resposta Transcrita do Aluno

> "Além de regularizar o modelo, se atentar a métrica f1-score que é usada quando um falso negativo custa caro, mas também um falso positivo, para ver se o modelo está equilibrado."

Parecer do Docente

O aluno sugere utilizar o F1-score como indicador equilibrado. No entanto, o subitem solicitava cautela amostral perante o volume escasso de imagens noturnas: era mandatório pontuar o perigo de conclusões precipitadas sem significância estatística, exigindo intervalos de confiança, bootstrap ou validação cruzada estratificada.

Questão 3 — Alerta que muda decisões (2,4 / 5,0 pts)

3.a Interpretação da Calibração e Crítica ao Limiar 0,50 (0,5 / 2,0)

Resposta Transcrita do Aluno

> "Pois não leva em conta os custos dos falsos positivos e falsos negativos, que fazem esse classificador quebrar, por isso ele está mal calibrado."

Parecer do Docente

Resposta superficial. O aluno não leu o diagrama de calibração (Figura 2), no qual a probabilidade predita de 0,50 corresponde na realidade a apenas 18% de casos positivos reais, evidenciando sobreconfiança drástica do modelo. Também não definiu calibração empírica (onde p_hat reflete a frequência real observada, reta y=x).

3.b Custos Assimétricos e Capacidade da Equipe Médica (1,4 / 2,0)

Resposta Transcrita do Aluno

> "Tem que avaliar qual o mais custoso, se falso positivo, usar a métrica precision se falso negativo, usar recall, se ambos, f1-score. Provavelmente, o recall é a mais adequada, visto que o objetivo é preservar a vida humana."

Parecer do Docente

Compreensão sólida da assimetria clínica de custos: o falso negativo (não detectar um paciente crítico) coloca vidas em risco, justificando a primazia do recall. Contudo, não ponderou o teto físico de atendimento da equipe médica (vazão de chamados), fundamental para equilibrar o limiar operacional e evitar a fadiga de alarmes falsos.

3.c Telemetria Contínua e Deriva de Conceito (0,5 / 1,0)

Resposta Transcrita do Aluno

> "Ter um calibrador automático, se a equipe médica reportar uma anomalia ele recalcula e ajusta, e aprende."

Parecer do Docente

Válido ao introduzir feedback da equipe médica para disparar recalibração. Faltou descrever um plano formal de telemetria em produção (acompanhamento contínuo de métricas diárias e detecção de concept drift) e ações técnicas de salvaguarda (como isolamento temporário e fallback para triagem humana direta).

Questão 4 — Busca de conhecimento para atendimento (3,4 / 5,0 pts)

4.a Representação Semântica e Limitações da Similaridade de Cosseno (1,3 / 2,0)

Resposta Transcrita do Aluno

> "Um contextualizado supera um estático pois verifica os vetores mais próximos para retornar. O cosseno não lida com todos por exemplo, pode retornar algo obsoleto, um documento irrelevante, ou não retornar nada se não tiver algo naquilo dito."

Parecer do Docente

Apontou com perspicácia que o cosseno isolado pode recuperar normas obsoletas e irrelevantes por proximidade temática. Contudo, a explicação de estático versus contextualizado foi rasa, não explicitando que o Word2Vec gera vetores estáticos únicos por palavra (ignorando sintaxe e polissemia), enquanto modelos baseados em Transformers geram representações dinâmicas condicionadas ao contexto da sentença.

4.b Avaliação Rigorosa de Recuperação (Information Retrieval) (0,9 / 1,5)

Resposta Transcrita do Aluno

> "Usar a proporção de consultas que retorna resultado é ruim pois se o que está sendo retornado não for útil como resposta é uma métrica vazia. É necessário um protocolo que melhore o retorno para algo relevante, com métricas melhores como: data de publicação (mais recente melhor)."

Parecer do Docente

Crítica muito bem colocada ao demonstrar que a mera proporção de retorno é uma métrica ingênua e vazia. Entretanto, não formalizou o protocolo de IR exigido pelo enunciado: construção de conjunto anotado de consultas com ground truth e aplicação de métricas canônicas de ranking (como Recall@K, Precision@K, MRR ou NDCG).

4.c Busca Híbrida e Governança Normativa (1,2 / 1,5)

Resposta Transcrita do Aluno

> "Adicionar uma busca léxica ajuda a aumentar a taxa de sucesso, pois além da busca semântica representada por vetores densos usando a similaridade cosseno, garante que a busca sintática encontre exatamente o que se procura, visto que ela analisa caracteres em ordem específica."

Parecer do Docente

Muito boa fundamentação da abordagem híbrida: reconheceu com precisão a complementaridade entre o espaço semântico denso e a correspondência léxica/sintática exata de termos (BM25). Faltou apenas explicitar os filtros de metadados estruturados (campo vigência/data) para garantir o descarte de normas revogadas.

Questão 5 — Assistente que pode não responder (3,7 / 5,0 pts)

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

Resposta Transcrita do Aluno

> "É possível que nem sempre o maior escore seja o documento esperado. Por isso, recomenda-se estratégias alternativas como retornar a partir de um limiar mínimo, ou comparar a diferença do primeiro pro segundo resultado, pois pode ser ilusória."

Parecer do Docente

Resposta brilhante e impecável! O aluno contestou o gestor e desenhou perfeitamente a política canônica de duplo limiar preconizada no gabarito: estabelecimento de um escore mínimo de confiança (tau_1) combinado com a exigência de uma margem de separação entre o primeiro e o segundo resultado (tau_2).

5.b Conflito de Fontes e Validação Prévia da Abstenção (1,3 / 2,0)

Resposta Transcrita do Aluno

> "Verificar a atualidade dos documentos, se as normas são conflituosas provavelmente a mais atual é mais válida. Para além disso, normas são superiores em hierarquia comparado com perguntas/consultas, logo sua resposta deve prevalecer. É possível validar testando."

Parecer do Docente

Correto na definição de hierarquia normativa e prevalência temporal da norma mais recente. No entanto, a proposta de validação prévia foi vaga ("validar testando"), não descrevendo a metodologia experimental essencial: testar o pipeline contra um benchmark balanceado com perguntas respondíveis e consultas intencionalmente ambíguas ou sem resposta no acervo.

5.c Telemetria da Taxa de Abstenção em Produção (0,4 / 1,0)

Resposta Transcrita do Aluno

> "Ela fornece a informação se o modelo generaliza ou não. É necessário ver como ele responde para mudanças para entender isso."

Parecer do Docente

Resposta genérica. O monitoramento contínuo da taxa de abstenção em sistemas RAG serve como sinal imediato para os engenheiros de dados: se a taxa sobe, indica deriva nas consultas dos usuários (query drift) ou o surgimento de lacunas documentais após revogações de normas, orientando a curadoria do acervo.

Laudo gerado oficialmente pela disciplina Inteligência Artificial II (IA II) • CEFET-MG Timóteo