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:** Danilo Fróis Kaizer
- **Matrícula:** 20233004010
- **Turma:** Turma 1
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 18,0 / 25,0** (Aproveitamento: **72,0%**)
- **Legibilidade da Caligrafia:** 92% (Alta Legibilidade)

Desempenho Consolidado por Questão

Questão / Tema	Pontos Obtidos	Valor	Aproveitamento
Q1: Triagem com dados heterogêneos	**4,6**	5,0	92,0%
Q2: Inspeção visual em condições novas	**4,1**	5,0	82,0%
Q3: Alerta que muda decisões	**3,2**	5,0	64,0%
Q4: Busca de conhecimento para atendimento	**3,5**	5,0	70,0%
Q5: Assistente que pode não responder	**2,6**	5,0	52,0%
TOTAL GERAL	**18,0**	**25,0**	**72,0%**

Questão 1 — Triagem com dados heterogêneos (4,6 / 5,0 pts)

1.a Diagnóstico da Curva de Aprendizado e Decisão de Produção (1,8 / 2,0)

Resposta Transcrita do Aluno

> "A partir da época 8 o modelo começa a dar overfitting, ou seja, começa a decorar padrões do conjunto de treino. Isso justifica a diminuição do loss (começa a acertar mais no treino) e o aumento do val_loss (perde a generalização e começa a errar). A afirmação da equipe está parcialmente correta: o modelo está pronto, mas não é pelo motivo dado, e sim porque o val_loss começou a subir, o que deixa claro que a época com os melhores pesos é a 8."

Parecer do Docente

Excelente diagnóstico do sobreajuste a partir da Época 8 e identificação precisa do ponto ótimo de parada (checkpoint de pesos). Pequena ressalva de que a afirmação da equipe técnica era conceitualmente falsa em sua totalidade (pois pretendiam manter o treinamento contínuo baseando-se apenas na perda de treino).

1.b Representação de Entradas Heterogêneas e Retropropagação (1,8 / 2,0)

Resposta Transcrita do Aluno

> "As entradas numéricas devem ser normalizadas e as categóricas devem ser convertidas em numéricas, por exemplo, via One Hot ou Ordinal Encoding. O papel do backpropagation é de calcular o erro dos neurônios de trás pra frente, usando uma abordagem recursiva e programação dinâmica. Esse erro é utilizado para corrigir os pesos do modelo durante o treinamento."

Parecer do Docente

Muito boa conceituação técnica, conectando o cálculo de gradientes por regra da cadeia com programação dinâmica/recursão e indicando normalização e One-Hot Encoding.

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

Resposta Transcrita do Aluno

> "É essencial dividir a base em conjuntos de treino (ajuste dos pesos), teste (conjunto independente para validar o final) e validação (evitar overfitting no treino). Além disso, para avaliar o modelo é imprescindível utilizar várias métricas em conjunto: acurácia, precisão, recall e f1-score. Para reprodutibilidade, deve-se utilizar uma seed específica para geração de números randômicos (Ex: 42)."

Parecer do Docente

Completo e preciso: descreveu a divisão das três partições, o conjunto diversificado de métricas e a fixação de semente aleatória para reprodutibilidade experimental.

Questão 2 — Inspeção visual em condições novas (4,1 / 5,0 pts)

2.a Diagnóstico Experimental do Aumento de Alertas (1,6 / 2,0)

Resposta Transcrita do Aluno

> "Para investigar, seria interessante fazer um teste de estresse do modelo com foto de peças no período noturno e matutino. Uma hipótese para o problema é de que, ao invés de aprender padrões de peças com defeito, o modelo esteja identificando os padrões de luzes nas fotos. Supondo que a noite tem mais peças defeituosas por motivos biológicos do corpo humano, o modelo pode ter aprendido que vale mais a pena mandar o alerta a noite (também pode ser causado por uma amostra ruim)."

Parecer do Docente

Boa percepção prática sobre correlação espúria e efeito da iluminação noturna (covariate shift). Faltou discriminar de forma independente as hipóteses de sobreajuste geral e inconsistência de rotulação humana.

2.b Arquitetura Convolucional e Transferência de Aprendizado (1,8 / 2,0)

Resposta Transcrita do Aluno

> "Para trabalhar com imagens, CNN é o melhor modelo, pois foi desenhado para isso. Usar MLP por exemplo iria explodir o número de parâmetros. Ao usar um modelo pré-treinado, o ideal é congelar o backbone nas primeiras épocas até que o modelo head se estabilize e, somente depois, descongelá-lo. Isso evita a destruição dos pesos pré-calculados e permite um ajuste fino ao final do modelo."

Parecer do Docente

Justificativa correta da CNN (evitar explosão de parâmetros) e protocolo impecável de transferência de aprendizado (congelamento do backbone com aquecimento da cabeça classificadora antes do descongelamento parcial). Pequena dedução pela ausência do diagrama visual.

2.c Cautela Amostral e Validação com Poucos Dados (0,7 / 1,0)

Resposta Transcrita do Aluno

> "Uma solução seria aumentar o número de imagens de peças à noite, por exemplo, tirando fotos à noite das peças produzidas de manhã ou aplicando filtros de imagem."

Parecer do Docente

Propôs aumento de dados e balanceamento de luminosidade, mas não abordou os procedimentos estatísticos formais exigidos para atestar superioridade (validação cruzada estratificada e teste de significância).

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

3.a Interpretação da Calibração e Crítica ao Limiar 0,50 (1,2 / 2,0)

Resposta Transcrita do Aluno

> "O limiar fixado do modelo depende do contexto do problema, no caso de alarmes falsos, o ideal seria aumentar o limiar para evitá-los. É justamente esse ajuste contextual que define a calibração do modelo. Mas é importante destacar que essa decisão é um trade-off: aumentar o limiar aumenta a precisão e diminui o recall."

Parecer do Docente

Discutiu corretamente o trade-off entre precisão e revocação ao deslocar o limiar. Contudo, não realizou a leitura analítica dos dados quantitativos da Figura 2 (18% real para $\hat{p} = 0{,}50$, resultando em 82% de alarmes falsos) nem distinguiu o conceito de calibração probabilística da simples escolha de limiar de corte.

3.b Custos Assimétricos e Capacidade da Equipe Médica (1,6 / 2,0)

Resposta Transcrita do Aluno

> "A equipe deve avaliar se o custo é maior com alarmes falsos, cuja solução é aumentar o limiar, ou se um falso negativo (paciente que precisa de ajuda mas o alarme não toca) é mais custoso, cuja solução é diminuir o limiar."

Parecer do Docente

Mapeou com clareza o efeito direcional do limiar sobre falsos positivos e negativos em cenários de custos assimétricos. Faltou explicitar a restrição de capacidade operacional da equipe e formalizar o custo esperado.

3.c Telemetria Contínua e Deriva de Conceito (0,4 / 1,0)

Resposta Transcrita do Aluno

> "Uma das formas de monitorar o treinamento do modelo é usar um conjunto de validação, o qual irá auxiliar o engenheiro a tomar a decisão correta."

Parecer do Docente

Resposta restrita ao processo de treino/validação offline. Não abordou telemetria em produção (monitoramento de escores em tempo real, detecção de drift ou planos de contingência).

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

4.a Representação Semântica e Limitações da Similaridade de Cosseno (1,5 / 2,0)

Resposta Transcrita do Aluno

> "O uso de embeddings estáticos vai limitar a busca a encontrar padrões léxicos na base. O uso de transformers permitiria a capacidade de generalização por similaridades semânticas."

Parecer do Docente

Distinguiu o limite de embeddings estáticos frente ao ganho semântico de Transformers. Deixou de pontuar as fraquezas específicas da similaridade de cosseno (insensibilidade a metadados e forçamento de top-k).

4.b Avaliação Rigorosa de Recuperação (Information Retrieval) (0,6 / 1,5)

Resposta Transcrita do Aluno

> "Um protocolo que soluciona esse problema é o RAG, pois ele avalia a proporção e o escore de consultas e decide se é suficiente para a LLM dar a resposta ou se há falta de informação na base."

Parecer do Docente

Mencionou RAG no contexto da tomada de decisão, mas não apresentou o protocolo de avaliação em Information Retrieval exigido pelo gabarito (consultas anotadas e métricas como MRR, NDCG e Recall@K).

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

Resposta Transcrita do Aluno

> "A busca híbrida combina a melhor parte das buscas lexical e semântica. Se o modelo encontrar uma correspondência direta, como um código, nome ou norma, a busca léxica toma a frente, caso contrário, a busca semântica e a generalização se impõem."

Parecer do Docente

Excelente raciocínio funcional sobre o papel de cada modalidade de busca (léxica para códigos e termos exatos, semântica para expansão conceitual). Faltou apenas citar a filtragem por metadados para controle de vigência.

Questão 5 — Assistente que pode não responder (2,6 / 5,0 pts)

5.a Contestação do Maior Escore e Desenho da Abstenção (1,7 / 2,0)

Resposta Transcrita do Aluno

> "A decisão do gestor é ruim, pois o documento de maior escore não é necessariamente o melhor. O correto seria utilizar um limiar de escore para o modelo caso não tenha evidência suficiente e outro para avaliar documentos com resultados menores que o primeiro."

Parecer do Docente

Correta contestação da regra ingênua e proposta conceitual válida de dois limiares (um de corte por falta de evidência e outro de margem comparativa entre os documentos).

5.b Conflito de Fontes e Validação Prévia da Abstenção (0,9 / 2,0)

Resposta Transcrita do Aluno

> "Uma das ideias seria pré computar algumas buscas e repostas. Assim, o modelo não entraria em conflito de normas antes de emitir seu resultado."

Parecer do Docente

Sugeriu pré-computação como controle prévio, mas não detalhou os critérios de desempate por hierarquia/vigência de normas nem o protocolo de validação com consultas propositalmente não respondíveis.

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

Resposta Transcrita do Aluno

> [Subitem deixado em branco pelo discente]

Parecer do Docente

Sem resposta apresentada. Pontuação não atribuída.

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