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:** Gabriel Duarte Campos Amorim
- **Matrícula:** 20223007378
- **Turma:** Turma 2
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 13,5 / 25,0** (Aproveitamento: **54,0%**)
- **Legibilidade da Caligrafia:** 86% (Média Legibilidade)

Desempenho Consolidado por Questão

Questão / Tema	Pontos Obtidos	Valor	Aproveitamento
Q1: Triagem com dados heterogêneos	**3,2**	5,0	64,0%
Q2: Inspeção visual em condições novas	**2,7**	5,0	54,0%
Q3: Alerta que muda decisões	**3,1**	5,0	62,0%
Q4: Busca de conhecimento para atendimento	**3,5**	5,0	70,0%
Q5: Assistente que pode não responder	**1,0**	5,0	20,0%
TOTAL GERAL	**13,5**	**25,0**	**54,0%**

Questão 1 — Triagem com dados heterogêneos (3,2 / 5,0 pts)

1.a Diagnóstico da Curva de Aprendizado e Decisão de Produção (1,6 / 2,0)

Resposta Transcrita do Aluno

> "O comportamento das curvas de perda revela que o modelo está sofrendo com overfitting. A queda contínua da perda de treino não garante que o sistema está pronto porque a queda da validação parou de cair o que indica que o modelo está apenas decorando os resultados da fase de treino e os replicando ao invés de aprender com eles."

Parecer do Docente

Boa explicação conceitual de sobreajuste (overfitting), diferenciando memorização de aprendizado generalizável. Faltou referenciar diretamente a Época 8 e indicar a política de parada antecipada (early stopping).

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

Resposta Transcrita do Aluno

> "As entradas heterogêneas devem ser representadas pelo uso de hiperparâmetros. A retropropagação é importante pois permite analisar e ajustar os pesos durante o período de treinamento, permitindo que o modelo consiga um melhor resultado."

Parecer do Docente

Resposta imprecisa no pré-processamento (hiperparâmetros não representam atributos heterogêneos; exigem-se padronização/Z-score e One-Hot/Embeddings). Na retropropagação, reconheceu o ajuste de pesos, mas omitiu a fundamentação dos gradientes via regra da cadeia.

1.c Evidências Clínicas de Confiança e Reprodutibilidade (0,8 / 1,0)

Resposta Transcrita do Aluno

> "Para confiar no modelo é necessário analisar dados independentes e também montar e analisar a matriz de confusão."

Parecer do Docente

Correta identificação da necessidade de partição independente de teste cego e análise da matriz de confusão. Faltou citar o controle de sementes aleatórias para reprodutibilidade.

Questão 2 — Inspeção visual em condições novas (2,7 / 5,0 pts)

2.a Diagnóstico Experimental do Aumento de Alertas (0,8 / 2,0)

Resposta Transcrita do Aluno

> "No começo o modelo estaria passando pela fase de teste e validação e por isso houve muitas taxas de alerta. Podemos utilizar o AUC como métrica para distinguir as hipóteses de sobreajuste, mudança de distribuição e falha no processo de rotulação."

Parecer do Docente

Mencionou a métrica AUC, mas não explicou como ela isolaria as três hipóteses concorrentes. Faltou detalhar o experimento comparativo direto entre imagens diurnas e noturnas (covariate shift) e a auditoria de consistência dos inspetores humanos.

2.b Arquitetura Convolucional e Transferência de Aprendizado (1,4 / 2,0)

Resposta Transcrita do Aluno

> "A CNN é apropriada porque podemos pegar apenas uma parte dela para usar como modelo e guiar os testes. O restante da base ficaria congelado (com seu todo) e o modelo que está sendo treinado poderia ser ajustado conforme fosse necessário. Durante o ajuste poderia haver um aumento dos dados analisados no modelo para combater a escassez."

Parecer do Docente

Compreendeu a estratégia de transferência de aprendizado (congelamento do backbone com ajuste de camadas superiores) e aumento de dados. Não detalhou as propriedades convolucionais espaciais nem forneceu o diagrama visual solicitado.

2.c Cautela Amostral e Validação com Poucos Dados (0,5 / 1,0)

Resposta Transcrita do Aluno

> "Para evitar conclusões precipitadas poderíamos trocar ou remover as imagens. Se o modelo estivesse tendo melhorias e após essa alteração ele apresentasse uma queda no desempenho, então encontraríamos como resposta uma dependência entre as imagens e as melhorias aplicadas e isso nos afirma que os resultados encontrados estão válidos."

Parecer do Docente

Propôs teste de dependência por modificação amostral. Faltou abordar as práticas metodológicas padrão: validação cruzada estratificada, teste cego específico no turno noturno e testes de significância estatística.

Questão 3 — Alerta que muda decisões (3,1 / 5,0 pts)

3.a Interpretação da Calibração e Crítica ao Limiar 0,50 (0,8 / 2,0)

Resposta Transcrita do Aluno

> "A premissa de fixar esse limiar é falha porque o valor estabelecido não condiz com a realidade e por isso houve muitos alarmes inúteis. Estar mal calibrado é justamente isso, incluir valores que não agregam em nada o nosso modelo."

Parecer do Docente

Identificou que o limiar 0,50 gera alarmes inúteis por descompasso com a realidade. Não realizou a leitura quantitativa do gráfico da Figura 2 (18% real para $\hat{p} = 0{,}50$, resultando em 82% de alarmes falsos) nem definiu o conceito formal de calibração de probabilidades.

3.b Custos Assimétricos e Capacidade da Equipe Médica (1,8 / 2,0)

Resposta Transcrita do Aluno

> "FP, FN e teto da equipe precisam ser mensurados para determinar o limiar. Se grande parte da equipe não atender os FN porque estiveram ocupados com os FP é grave. É preciso haver pesos em cada uma dessas variáveis, pois a equipe atender FP é aceitável, mas negligenciar FN é gravíssimo."

Parecer do Docente

Excelente articulação contextual! Ponderou com muita clareza que negligenciar FN é inaceitável, enquanto atender FP é tolerável desde que não sature o teto da capacidade da equipe médica.

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

Resposta Transcrita do Aluno

> "Observar queda na curva de validação. Uma anomalia precisa ser tratada imediatamente e deve-se adicionar ao modelo o que foi aprendido com ela"

Parecer do Docente

Focou no acompanhamento em validação. Deixou de detalhar a telemetria em tempo real em produção (distribuição de escores e taxas de alarmes) e plano de fallback seguro.

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

4.a Representação Semântica e Limitações da Similaridade de Cosseno (1,6 / 2,0)

Resposta Transcrita do Aluno

> "Contextualizados são mais rápidos e usarão a palavra de contexto para prever a palavra central enquanto estáticos usam a palavra central para prever o contexto A limitação é que o cosseno vai medir o alinhamento vetorial dos documentos recuperados e não vai representar uma probabilidade de relevância do que foi procurado."

Parecer do Docente

Brilhante reflexão sobre a limitação do cosseno: mede apenas alinhamento vetorial geométrico, o que não equivale a uma probabilidade real de relevância da norma. Houve apenas imprecisão na distinção CBOW/Skip-gram vs Transformers.

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

Resposta Transcrita do Aluno

> "É inadequado medir o sucesso apenas pela proporção de consultas com retorno porque não está sendo analisado as buscas que não retornaram nada e nem investigando se há padrão entre elas."

Parecer do Docente

Percebeu a fraqueza da métrica superficial de retorno. Faltou estruturar o protocolo com consultas de teste anotadas e métricas consagradas de Information Retrieval (MRR, NDCG e Recall@K).

4.c Busca Híbrida e Governança Normativa (1,3 / 1,5)

Resposta Transcrita do Aluno

> "Uma abordagem híbrida mesclaria a busca semântica com a busca léxica e essa combinação supera a busca vetorial pura por apresentar uma maior quantidade de fatores (exemplo: termos, identificadores) válidos, fazendo com que seja possível superar as limitações e encontrar um bom resultado."

Parecer do Docente

Excelente justificativa da complementaridade entre busca semântica e léxica (identificadores e termos exatos). Faltou apenas explicitar os filtros por metadados para controle de vigência documental.

Questão 5 — Assistente que pode não responder (1,0 / 5,0 pts)

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

Resposta Transcrita do Aluno

> "A proposta do gestor de fazer com que o sistema sempre mostre o documento de maior escore pois isso elimina a necessidade de regras adicionais é VÁLIDO porque consultas ambíguas, documentos conflitantes e mudanças frequentes nas normas são comuns fazendo com que as respostas automáticas sejam geradas com base nesses fatores."

Parecer do Docente

Conclusão oposta à diretriz de engenharia. O gabarito exige a contestação da proposta do gestor, demonstrando que forçar sempre o maior escore induz alucinações e respostas incorretas em perguntas fora de domínio ou ambíguas.

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

Resposta Transcrita do Aluno

> "É possível tratar documentos com normas conflitantes através de embedding. Podemos validar a abstenção por meio da similaridade do cosseno."

Parecer do Docente

Muito superficial. Similaridade vetorial pura não resolve conflito normativo (que exige hierarquia e recência de metadados) nem valida abstenção sem um dataset com perguntas não respondíveis.

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

Resposta Transcrita do Aluno

> "Podemos utilizar um early stopping definindo uma paciência X para determinar se houve uma mudança na taxa de abstenção durante um período X de épocas e emita um sinal caso ocorra uma mudança. Esse sinal fornecerá aos engenheiros a informação de um uso até os dias atuais e quanto tempo o modelo fica sem alterações."

Parecer do Docente

Criou uma metáfora interessante com early stopping para alertar sobre mudanças na taxa de abstenção. Faltou conectar esse indicador com alterações práticas de perfil de usuário ou obsolescência da base regulatória.
