# 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:** Iviner Cássio de Matos Santos
- **Matrícula:** 20233004370
- **Turma:** Turma 1
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 22,3 / 25,0** (Aproveitamento: **89,2%**)
- **Legibilidade da Caligrafia:** 94% (Alta Legibilidade)

### Desempenho Consolidado por Questão

| Questão / Tema | Pontos Obtidos | Valor | Aproveitamento |
|---|---|---|---|
| **Q1:** Triagem com dados heterogêneos | **4,7** | 5,0 | 94,0% |
| **Q2:** Inspeção visual em condições novas | **4,0** | 5,0 | 80,0% |
| **Q3:** Alerta que muda decisões | **4,3** | 5,0 | 86,0% |
| **Q4:** Busca de conhecimento para atendimento | **4,5** | 5,0 | 90,0% |
| **Q5:** Assistente que pode não responder | **4,8** | 5,0 | 96,0% |
| **TOTAL GERAL** | **22,3** | **25,0** | **89,2%** |

## Questão 1 — Triagem com dados heterogêneos (4,7 / 5,0 pts)

### 1.a Diagnóstico da Curva de Aprendizado e Decisão de Produção (1,9 / 2,0)

Resposta Transcrita do Aluno

> "As curvas, a partir da época 08, indicam que o modelo deve estar em overfitting. Ou seja, que ele decorou os dados de treino e perdeu poder de generalização para novos dados. Por isso a perda de treino segue diminuindo, e a perda de avaliação (dados desconhecidos para ajuste) segue aumentando."

Parecer do Docente

Diagnóstico excelente: apontou a bifurcação das curvas a partir da Época 8, explicou o sobreajuste como perda de capacidade de generalização e vetou a ida para produção.

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

Resposta Transcrita do Aluno

> "Os dados devem ser pré-processados antes de serem inseridos no modelo MLP, mantendo os valores numéricos e aplicando, por exemplo, one-hot encoding nos categóricos. Eles devem ser representados como vetores, matrizes ou tensores. No backpropagation, os parâmetros do modelo são recalculados, de acordo com o feedback obtido com os dados de validação no treinamento. Este mecanismo que penaliza e ajusta os parâmetros conformes os erros."

Parecer do Docente

Muito boa conceituação de representação vetorial e One-Hot Encoding, além do ajuste iterativo de pesos pelo backpropagation. Faltou apenas especificar o escalonamento/normalização das variáveis numéricas.

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

Resposta Transcrita do Aluno

> "Antes de colocar o modelo em produção, é importante avaliar o seu desempenho com dados novos independentes (conjunto de teste). Além de considerar métricas com recall, precision e F1-score, juntamente com a acurácia. Para que os experimentos sejam reprodutíveis, é importante esta divisão de base (treino/validação/teste) estratificada e com uma semente aleatória fixada."

Parecer do Docente

Pontuação máxima. Cobriu o protocolo completo: partição independente de teste, métricas clínicas adequadas (Recall, Precisão, F1-Score) e garantia de reprodutibilidade via divisão estratificada com fixação de semente aleatória.

---

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

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

Resposta Transcrita do Aluno

> "Na investigação, deve verificar se houve divisão e estratificação da base de dados, já que ela é desbalanceada. Verificar se o modelo teve um aprendizado 'raso', por exemplo, considerando a claridade ao invés da peça na imagem (sem dropout). Nestas verificações, deve-se expor o modelo a uma base de teste de confiança, ou rotulada a mão, e avaliar suas métricas de desempenho."

Parecer do Docente

Excelente percepção do atalho espúrio decorrente da claridade/iluminação (covariate shift) e proposta de auditoria com base rotulada manualmente. Faltou explicitar a verificação de sobreajuste geral.

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

Resposta Transcrita do Aluno

> "Uma rede CNN é o modelo recomendado para classificação e reconhecimento de padrões em imagens. Ela permite tal processamento com menos parâmetros (economia computacional), utilizando a técnica de análise com kernels. → Neste cenário, deve-se aplicar um congelamento nos parâmetros pré-treinados, e um data augmentation para simular novas imagens e melhorar o desbalanceamento da base. Por fim, treinar um MLP de 'cabeça', especializado neste reconhecimento, que é alimentado pelos parâmetros de saída do 1º modelo."

Parecer do Docente

Domínio exemplar dos fundamentos convolucionais (economia de parâmetros com kernels locais), congelamento de base pré-treinada, data augmentation e fine-tuning na cabeça MLP. Pequeno desconto apenas por não ter esquematizado o diagrama gráfico.

### 2.c Cautela Amostral e Validação com Poucos Dados (0,6 / 1,0)

Resposta Transcrita do Aluno

> "No modelo MLP de 'cabeça', deve-se configurar uma rede com dropout, para induzir um aprendizado profundo entre os neurônios. Isso evita dependência entre os neurônio, aumenta a absorção de características do modelo. (dropout durante o treinamento)."

Parecer do Docente

Abordou regularização por dropout para evitar coadaptação de neurônios. Contudo, não tratou dos procedimentos estatísticos de validação exigidos no subitem (validação cruzada estratificada e teste de significância estatística).

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

### 3.a Interpretação da Calibração e Crítica ao Limiar 0,50 (1,8 / 2,0)

Resposta Transcrita do Aluno

> "O limiar de 0,5 está incorreto, uma vez que a frequência real atinge o percentual de 0,5 em $\hat{p} = 0,9$, aproximadamente. Neste caso, o modelo está alertando a equipe médica que o paciente terá uma piora e precisa de intervenção rápida deles, entretanto, muitas vezes erradando. Para $\hat{p} = 0,5$, errando mais do que acertando. Isso pode gerar exames extras, transtorno e estresse desnecessário."

Parecer do Docente

Excelente leitura quantitativa da Figura 2: demonstrou que no limiar 0,50 o modelo 'erra mais do que acerta' (taxa real 18%), denunciando estresse desnecessário e sobrecarga assistencial gerada por má calibração.

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

Resposta Transcrita do Aluno

> "Os custos de um falso negativo do modelo, podem ser maiores que os de um falso positivo, uma vez que no primeiro, um paciente que precisava de atenção rápida pode não ser atendido. Usando esse critério, deve-se preferir um limiar 'precavido', que não deixe de identificar os verdadeiros positivos, mas respeite o número de pacientes e médicos disponíveis na clínica/hospital. Evita gastos desnecessários, sobrecarga médica, e não atendimentos."

Parecer do Docente

Resposta exemplar! Alinhou com maestria a urgência de evitar falsos negativos (risco de morte) com a restrição operacional de médicos disponíveis, impedindo colapso assistencial por excesso de alarmes.

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

Resposta Transcrita do Aluno

> "Considerar retreinar o modelo com o critério de otimizar os verdadeiros positivos e minimizar os falsos negativos."

Parecer do Docente

Citou retreino com otimização assimétrica. Faltou detalhar a telemetria em produção (acompanhamento contínuo de escores e taxas de alarmes) e os planos de fallback hospitalar.

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

### 4.a Representação Semântica e Limitações da Similaridade de Cosseno (1,7 / 2,0)

Resposta Transcrita do Aluno

> "A similaridade de cosseno metrifica o quanto um conceito (palavra) está próxima de outra, baseado no contexto que elas foram apresentadas durante o treinamento. Entretanto, mesmo considerando a base de treino como validada, palavras que compartilham contexto aparecem próximos, mesmo que não sejam sinônimas. Logo, outras métricas e avaliações também são necessárias."

Parecer do Docente

Arguta reflexão sobre o ponto cego da similaridade semântica por cosseno: proximidade contextual de coocorrência não garante equivalência semântica ou relevância legal da norma recuperada.

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

Resposta Transcrita do Aluno

> "Medir apenas similaridade de cosseno não é adequado, palavras com contextos próximos têm alta similaridade, mesmo não tendo significados equivalentes. Para melhor a busca, pode usar RAG, busca semântica e léxica, e avaliar o Recall@K e Precision@K, por exemplo."

Parecer do Docente

Muito bom. Citou expressamente o paradigma RAG e a avaliação formal com métricas clássicas de Information Retrieval (Recall@K e Precision@K).

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

Resposta Transcrita do Aluno

> "A combinação destas técnicas de busca, tendem a melhorar as respostas, já que analisa-se o significado, contexto e a aparição da palavra nos vetores de retorno. Além, de analisar outros métricos com Recall@K e Precision@K, das K primeiras respostas. Todas trabalham para mitigar as limitações entre si, e localizam a melhor resposta num RAG."

Parecer do Docente

Articulou com rigor a complementaridade funcional entre busca léxica (palavras exatas) e semântica contextual. Faltou explicitar o mecanismo de filtros por metadados de vigência.

---

## Questão 5 — Assistente que pode não responder (4,8 / 5,0 pts)

### 5.a Contestação do Maior Escore e Desenho da Abstenção (1,8 / 2,0)

Resposta Transcrita do Aluno

> "Basear-se apenas no maior score não é o mais recomendado. Sendo, indicado ter um limiar para considerar uma resposta como válida ($\tau_1$), e um limiar ($\tau_2$) para envio de uma segunda opção de resposta. E neste último, ficando ao usuário avaliar qual a melhor resposta para ele."

Parecer do Docente

Correta contestação da regra do maior escore e formulação adequada dos dois limiares de decisão ( para corte de confiança mínima e para desambiguação inter-documentos).

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

Resposta Transcrita do Aluno

> "Para consultas com documentos conflitantes, pode-se optar para que o modelo responda baseado no documento com upload mais recente, a menos que a pergunta se refira explicitamente a datas anteriores. Usando um conjunto de documentos e consultas de teste, é possível simular e avaliar o funcionamento do modelo, tanto com consultas precisas (respondíveis) e não precisas (não respondíveis)."

Parecer do Docente

Excelente! Governança documental fundamentada em metadados de upload mais recente aliada a um protocolo experimental robusto com consultas propositalmente respondíveis e não respondíveis.

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

Resposta Transcrita do Aluno

> "Pode-se definir uma resposta padrão para consultas não respondíveis e contabilizá-las, dado na proporção total de consultas. Caso o percentual seja elevado, pode-se considerar mudar uma investigação: se os documentos estão desatualizados, se as perguntas estão mal formuladas, ou se deve-se mudar os critérios de resposta."

Parecer do Docente

Perfeito! Desenhou a telemetria com base na proporção de respostas de abstenção e interpretou os picos como sinalizadores de desatualização documental ou mudança no perfil de formulação das consultas.

---