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:** Luiz Henrique Santos Dias
- **Matrícula:** 20233004280
- **Turma:** Turma 1
- **Data da Avaliação:** 19/10/2026
- **Pontuação Final: 12,6 / 25,0** (Aproveitamento: **50,4%**)
- **Legibilidade da Caligrafia:** 93% (Alta 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	**2,4**	5,0	48,0%
Q3: Alerta que muda decisões	**2,1**	5,0	42,0%
Q4: Busca de conhecimento para atendimento	**2,1**	5,0	42,0%
Q5: Assistente que pode não responder	**2,8**	5,0	56,0%
TOTAL GERAL	**12,6**	**25,0**	**50,4%**

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

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

Resposta Transcrita do Aluno

> "A afirmação da equipe não é adequada, porque leva em consideração somente a diminuição dos erros nos dados de treinamento, sem levar em consideração o aumento dos erros nos dados de validação a partir da época 8. O gráfico da Figura 1 mostra que está acontecendo overfitting, ou seja, está ocorrendo memorização dos dados de treinamento. É necessário que sejam adotadas técnicas de regularização, Dropout para avaliar o modelo."

Parecer do Docente

Excelente refutação: apontou com precisão a bifurcação na Época 8 com ascensão do erro de validação como indicativo de sobreajuste/memorização e prescreveu técnicas de regularização (Dropout). Faltou apenas sugerir formalmente early stopping no ponto ótimo de perda mínima.

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

Resposta Transcrita do Aluno

> "As entradas heterogêneas devem ser representadas respectivamente cada uma com o seu próprio peso, evitando ao máximo introduzir viés para não prejudicar a análise. Numéricos e categóricos representam classes distintas, então é essencial que seja aplicada de forma correta a normalização dos dados, garantindo que tudo esteja na mesma escala."

Parecer do Docente

Identificou a necessidade fundamental de normalização/escalonamento para equiparar as grandezas numéricas. Omitiu, contudo, a codificação para atributos categóricos (One-Hot Encoding ou Embeddings densos) e o papel do algoritmo de backpropagation no cálculo dos gradientes via regra da cadeia.

1.c Evidências Clínicas de Confiança e Reprodutibilidade (0,6 / 1,0)

Resposta Transcrita do Aluno

> "Para poder confiar no modelo várias evidências adicionais são indispensáveis, tais como: Número de épocas; Matriz de confusão; Quais métricas (Precision, recall) foram utilizadas para garantir que o modelo seja confiável etc. No caso de um tratamento de câncer por exemplo, a acurácia não é a métrica mais adequada, o Recall (Penaliza falsos negativos) consegue passar mais credibilidade."

Parecer do Docente

Muito boa contextualização clínica sobre a superioridade do Recall para penalizar falsos negativos em triagem médica e uso da matriz de confusão. Faltou referir-se à partição independente de

teste cego e ao controle de reprodutibilidade por fixação de sementes aleatórias.

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

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

Resposta Transcrita do Aluno

> "Para investigar corretamente é necessário ter o conhecimento de quais atributos/variáveis estão sendo utilizados e qual é a quantidade de dados que está sendo usada para realizar a identificação dos defeitos nas imagens. Além disso, investigar o número de épocas utilizadas. Se o modelo apresentar pequeno erro nos dados de treinamento e um erro maior nos dados de validação é um sinal de overfitting. É interessante também entender quantos dados foram utilizados para treino, validação e teste."

Parecer do Docente

Destacou a auditoria do erro treino vs. validação para diagnosticar sobreajuste. Não estruturou, porém, a distinção nítida entre covariate shift (distribuição de luminância dia/noite) e ruído de rotulação humana (inconsistência entre anotadores).

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

Resposta Transcrita do Aluno

> "O CNN é a forma mais adequada e utilizada para trabalhar com imagens. Porque ele consegue lidar com as características das imagens, tais como: largura, altura, quantidade de pixels etc. Se o modelo utilizou uma taxa de aprendizado de 0,01, o ideal é que nesse ajuste fino seja utilizado um valor menor e ir descongelando aos poucos para verificar se está melhorando ou piorando o modelo."

Parecer do Docente

Boa conceituação sobre fine-tuning: propôs redução da taxa de aprendizado e descongelamento progressivo de camadas. Faltou fundamentar a invariância espacial/extração de bordas das CNNs e o uso de data augmentation com variações de brilho.

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

Resposta Transcrita do Aluno

> "As conclusões precipitadas podem ser evitadas analisando o resultado final e não criando cenários genéricos, ou seja, se conseguir identificar bem o defeito da imagem, não generalizar que em qualquer período ela vai conseguir identificar bem."

Parecer do Docente

Percebeu a falácia da generalização precipitada entre turnos. Não formalizou, no entanto, o procedimento metodológico de amostragem (validação cruzada estratificada por turno e testes estatísticos de significância com intervalos de confiança).

Questão 3 — Alerta que muda decisões (2,1 / 5,0 pts)

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

Resposta Transcrita do Aluno

> "A premissa da direção para fixar o limiar em 0,50 é falha porque após 0,50 os dados não são representados de forma correta. A calibração deve ser maior porque para um contexto clínico alarmes pouco úteis ou falsos têm um custo muito alto dependendo da situação."

Parecer do Docente

Criticou a fixação do limiar em 0,50 sob descalibração e apontou o custo clínico dos alarmes falsos. Omitiu a leitura quantitativa da curva da Figura 2 (onde $\hat{p} = 0{,}50$ equivale a apenas 18% de eventos reais, implicando 82% de alarmes espúrios).

3.b Custos Assimétricos e Capacidade da Equipe Médica (0,7 / 2,0)

Resposta Transcrita do Aluno

> "A combinação desses custos deve ser a menor possível de preferência de modo que no modelo final haja um equilíbrio, ou seja, todos os custos tenham pesos próximos, de modo que uma variável/atributo não sobreponha os demais. E a partir disso verificar se atende ao ponto operacional (limiar)"

Parecer do Docente

Mencionou a busca pelo ponto operacional de menor custo, mas não articulou a assimetria crítica do custo de omissão (falso negativo) versus o teto da capacidade física de atendimento da equipe médica.

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

Resposta Transcrita do Aluno

> "Após ser realizado o pré-processamento ou já ter um modelo pré-treinado pegar esse modelo e fazer pequenas repartições de modo que não haja prejuízo dos dados que foram utilizados, parâmetros etc e verificar cenários específicos com o intuito de tornar o modelo o mais adequado. E a ação técnica imediata que deve ser tomada é descobrir quais fatores (variáveis/atributos) levou a aquele cenário e utilizar métricas

de avaliações rigorosas.”

Parecer do Docente

Sugeriu investigação de fatores de causa-raiz ao notar degradação. Faltou especificar métricas de telemetria em produção (detecção de concept/data drift) e um protocolo de contingência hospitalar seguro com fallback.

Questão 4 — Busca de conhecimento para atendimento (2,1 / 5,0 pts)

4.a Representação Semântica e Limitações da Similaridade de Cosseno (0,8 / 2,0)

Resposta Transcrita do Aluno

> “A similaridade do cosseno é útil quando os dados, as variáveis tem correlações entre si. Nesse cenário está lidando com variáveis que possuem características bem distintas entre si, não conseguindo assegurar a relevância real de um documento.”

Parecer do Docente

Pontuou a insuficiência da similaridade do cosseno puro para assegurar relevância real. Não contrastou, porém, a natureza de modelos estáticos (Word2Vec) com modelos contextuais profundos (Transformers) nem mencionou a vigência jurídica.

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

Resposta Transcrita do Aluno

> “A acurácia não é o modelo mais adequado para ser utilizado como métrica para avaliar o desempenho de um modelo, com dados desbalanceados. Nesse cenário consulta com retorno pode ter mais dados do que consultas sem retorno, passando a impressão de que o modelo é bom, generaliza bem, sendo que na verdade pode alavancar prejuízos”

Parecer do Docente

Rejeitou com razão a contagem bruta de retorno como atestado de qualidade. Omitiu a menção a uma base de teste com gabarito de especialistas (ground-truth) e métricas consagradas de ranking em IR (MRR, nDCG, Recall@K).

4.c Busca Híbrida e Governança Normativa (0,7 / 1,5)

Resposta Transcrita do Aluno

> “A abordagem híbrida consegue lidar com dados que possuem características bem distintas, além de utilizar ferramentas que permitem automatizar o processo e devolver resultados mais confiáveis e precisos. Esta reduz tempo computacional, consumo de memória. Essa abordagem utiliza tecnologias sofisticadas a qual consigo compensar

por meio de passos lógicos e coerentes para identificar normas antigas e revogadas."

Parecer do Docente

Percebeu a necessidade de lógica para expurgo de normas revogadas. Não explicitou a composição arquitetural clássica da busca híbrida: fusão de vetores densos (semânticos) com busca lexical exata (BM25) e metadados de vigência.

Questão 5 — Assistente que pode não responder (2,8 / 5,0 pts)

5.a Contestação do Maior Escore e Desenho da Abstenção (1,4 / 2,0)

Resposta Transcrita do Aluno

"Contesto a proposta do gestor, porque maior score não significa que o modelo seja 100% confiável. Defino um limiar mínimo de x, se o documento passar de x há evidências de que ele conseguiu lidar com os casos que não comuns, porém essa é a 1ª Verificação. A 2ª Verificação ( ) terá o acréscimo de alguns cenários que seria a combinação dos fatos comuns para evidenciar se o modelo está adequado ou não."

Parecer do Docente

Excelente contestação teórica da regra ingênua de maior escore. Propôs política em dois estágios: limiar mínimo de confiança ( ) e verificação secundária ( ). Faltou apenas formular explicitamente como margem de separação entre o 1º e o 2º classificados.

5.b Conflito de Fontes e Validação Prévia da Abstenção (0,7 / 2,0)

Resposta Transcrita do Aluno

"O ideal é pegar essas normas conflitantes e verificar qual está sendo o conflito e armazená-lo em um atributo ou variável e analisar se esse conflito possui importância ou não para o modelo, caso a resposta seja sim ir fazendo ajustes de parâmetros, ajustes finos de modo a conseguir eliminar o conflito e devolver a norma OK"

Parecer do Docente

Sugeriu isolamento e análise dos conflitos em variáveis de controle. Omitiu o uso de critérios de hierarquia e vigência temporal nos metadados, além da elaboração de um benchmark prévio com perguntas deliberadamente sem resposta.

5.c Telemetria da Taxa de Abstenção em Produção (0,7 / 1,0)

Resposta Transcrita do Aluno

> "Score. Porque definindo os parâmetros e deixando que somente aqueles que tem pontuação mínima tenham uma resposta automática, enquanto que aqueles que não alcançarem são indícios de taxa de abstenção em produção. Ideal também que seja criado variáveis que emitem sinal quando acontecer alguma modificação."

Parecer do Docente

Boa percepção operacional: relacionou o escore mínimo diretamente ao controle da taxa de abstenção em produção e monitoramento de modificações no acervo. Faltou discutir o impacto do excesso de abstenções sobre a equipe humana de retaguarda.

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