



DESEMPENHO DA INTELIGÊNCIA ARTIFICIAL (IA) NA CORREÇÃO DE RELATÓRIOS ACADÊMICOS EM CURSOS DE ENGENHARIA

DOI: 10.37702/2175-957X.COBENGE.2025.6257

Autores: ÉSOLY MADELEINE BENTO DOS SANTOS, DENISE HIRAYAMA, TIAGO BRANDAO COSTA, TATIANA CANEDA SALAZAR RIBEIRO

Resumo: A Inteligência Artificial (IA) generativa tem ganhado destaque na educação, especialmente o ChatGPT na automação de avaliações acadêmicas. Apesar de seu potencial em agilizar e padronizar correções, ainda enfrenta limitações relacionadas à precisão e vieses. Este trabalho analisa o desempenho da IA na correção de relatórios acadêmicos em cursos de Engenharia da Universidade Federal Fluminense. Os relatórios das disciplinas de processamento de materiais cerâmicos e poliméricos foram avaliados em três modalidades: professor, livre, com critério. As notas atribuídas pela IA foram comparadas às dos docentes por meio de análises quantitativas e qualitativas. Os resultados indicam que, embora a IA apresente agilidade e forneça feedback estruturado, há discrepâncias significativas nas notas, principalmente em trabalhos com desempenho inferior. Conclui-se que, com orientação adequada, a IA pode ser uma ferramenta de apoio útil, mas não substitui a avaliação humana.

Palavras-chave: inteligência Artificial, ChatGPT, avaliação automatizada, engenharia, avaliação de relatórios

DESEMPENHO DA INTELIGÊNCIA ARTIFICIAL (IA) NA CORREÇÃO DE RELATÓRIOS ACADÊMICOS EM CURSOS DE ENGENHARIA

1 INTRODUÇÃO

A Inteligência Artificial (IA) generativa é um tipo de inteligência capaz de criar textos, músicas, imagens e códigos de programação (MCCARTHY *et al.*, 1955; NGUYEN *et al.*, 2022). Embora a discussão sobre a utilização da Inteligência Artificial (IA) em diversos campos do conhecimento tenha se intensificado recentemente, especialmente a partir de novembro de 2022 com o surgimento de *chatbots* baseados em modelos de linguagem de grande porte (LLM), como ChatGPT, Gemini e Meta.ai, a divulgação da IA remonta à década de 1950 (MCCARTHY *et al.*, 1955; NGUYEN *et al.*, 2022)).

Entre 1970 e 1980 surgiram os Sistemas Inteligentes de Ensino (ITS), que ofereciam lições personalizadas ajustando materiais e ritmo de acordo com o perfil do estudante. (CASTRO; MACIEL, 2024). O desenvolvimento de sistemas especialistas possibilitou IA identificar lacunas no aprendizado e fornecer feedback imediato aos estudantes e professores, representando um passo importante na personalização do aprendizado (CASTRO; MACIEL, 2024). Recentemente intensificaram-se os debates sobre o uso ético e regulatório da IA. A UNESCO dedicou o Dia Internacional da Educação de 2025 a Inteligência Artificial nesta oportunidade fez um apelo para que os estados membros invistam na formação de professores para o uso responsável dessa tecnologia no campo educacional (UNESCO, 2024).

A ascensão da IAGen impulsionou várias possibilidades de utilização por docentes. Entre elas, destacam-se: a automação de tarefas administrativas, a geração de ideias e conteúdos, apoio ao ensino e à aprendizagem, auxiliando professores na compreensão e aprimoramento de suas metodologias, e como assistente conversacional no suporte às atividades docentes (ALMEIDA *et al.*, 2024, BOULAY, 2023). Com o avanço do uso dessas tecnologias, pesquisadores têm se dedicado a explorar sua utilização como ferramenta de apoio para docentes (PIMENTEL *et al.*, 2023). Nesse contexto, um dos campos que mais desperta interesse é o da avaliação de atividades acadêmicas.

A avaliação de projetos e trabalhos, tradicionalmente, baseia-se no conhecimento e experiência do docente, em critérios claros e objetivos pré-definidos, bem como em princípios de justiça e transparência, visando assegurar coerência e imparcialidade. No entanto, esse processo ainda enfrenta desafios significativos. Um dos principais obstáculos é a complexidade do processo avaliativo, que exige um tempo considerável do professor, especialmente em turmas numerosas, tornando a atividade ainda mais exaustiva (KEUNING, HEEREN e JEURING, 2019). Além disso, manter objetividade e consistência ao longo do tempo pode ser desafiador, uma vez que fatores como desatenção, fadiga e vieses inconscientes podem influenciar as avaliações (JUKIEWICZ, 2024). De acordo com Castelblanco e colaboradores (2024), que compararam as notas atribuídas por avaliadores humanos e pelo ChatGPT na avaliação de projetos, observou-se que as discrepâncias eram mais acentuadas em projetos com pontuações mais baixas atribuídas por humanos. Isso evidencia os desafios das inteligências artificiais na identificação de desempenhos insatisfatórios em tarefas argumentativas complexas.

Ao utilizar IA para a avaliação de trabalhos de programação, Jukiewicz (2024) constatou, por meio do ChatGPT, que o tempo médio necessário para analisar e fornecer feedback para uma única consulta é de aproximadamente 10 segundos. O ChatGPT é capaz de classificar um trabalho como correto, parcialmente correto ou incorreto. Além disso, a IA não é afetada pela fadiga, sendo imparcial e livre de vieses em suas avaliações, o que reduz a ocorrência de erros na atribuição de notas (JUKIEWICZ, 2024). Por outro lado, Castelblanco e colaboradores (2024) entendem que esta questão do viés é altamente dependente da qualidade e dos dados nos quais o ChatGPT foi treinado, se os dados forem tendenciosos, os modelos herdarão esses vieses.

Ao usar o ChatGPT para avaliações, este pode fornecer feedback personalizado, oferecendo insights e explicações valiosas que contribuem significativamente para o aprimoramento e desenvolvimento de projetos e trabalhos (CASTELBLANCO, CRUZ-CASTRO, YANG, 2024). Os aspectos negativos da utilização do ChatGPT para a correção de trabalhos acadêmicos começam pelo acesso. Para utilizar a versão mais recente, com todas as suas ferramentas e funcionalidades disponíveis, é necessário aderir a um serviço pago. Além disso, em alguns casos, o modelo pode identificar erros inexistentes na solução proposta pelo aluno. Outra limitação é que a IA nem sempre retorna respostas idênticas para consultas semelhantes (JUKIEWICZ, 2024). Caso a avaliação seja realizada virtualmente apenas com as respostas construídas pelo ChatGPT, isso pode comprometer a interação entre alunos e professores, um aspecto fundamental para o desenvolvimento educacional. A troca de conhecimentos, o esclarecimento de dúvidas e o acompanhamento personalizado são elementos enriquecedores da aprendizagem que podem ser reduzidos com a automação do processo avaliativo (CASTELBLANCO, CRUZ-CASTRO, YANG, 2024).

O objetivo deste trabalho é analisar a eficiência da utilização da inteligência artificial como ferramenta de apoio na correção de relatórios realizados por estudantes dos cursos de Engenharia Metalúrgica e de Engenharia de Materiais da Universidade Federal Fluminense. Para isso, será avaliada a concordância entre as notas atribuídas por sistemas de IA e aquelas atribuídas por professores em diferentes disciplinas.

2 METODOLOGIA

2.1 Contextualização do estudo

O presente trabalho foi realizado por duas docentes da Escola de Engenharia Industrial Metalúrgica de Volta Redonda (EEIMVR) da Universidade Federal Fluminense (UFF) como objetivo de avaliar a eficácia da correção automática de avaliações de relatórios utilizando o modelo de linguagem ChatGPT, comparando os resultados com correções realizadas por professores. Foram avaliadas atividades coletadas de anos anteriores previamente corrigida, de duas disciplinas do Curso de Engenharia de Materiais:

- 1) Processamento de Materiais cerâmicos, do segundo semestre de 2024. A disciplina utiliza a metodologia de Aprendizagem Baseada em Projeto (PBL) como estratégia central de ensino. Os alunos produziram diversos relatórios técnicos que são parte da avaliação no contexto da utilização de PBL.
- 2) Processamento de termoplásticos. Os relatórios produzidos pelos alunos são resultado da aplicação da metodologia de Aprendizagem Baseada em Projeto (PBL) entre os anos de 2023 e 2024. Ao final do semestre, os grupos de alunos deveriam entregar um relatório consolidando todo o trabalho desenvolvido ao longo do período.

2.2 Atividades analisadas

Foram analisados relatórios técnicos das disciplinas de Processamento de Materiais cerâmico e Processamento de termoplásticos. As atividades desenvolvidas pelos alunos

seguiram uma estrutura de escrita previamente definida pelo professor, com tópicos e orientações específicas para a elaboração dos relatórios. Conforme a sistema de avaliação da disciplina foram atribuídas notas com base em uma escala numérica de 0 (zero) a 10 (dez).

2.3 Ferramentas de IA utilizadas

Foi utilizado o modelo de linguagem ChatGPT versão (GPT-4) disponível publicamente por meio da plataforma da OpenAI, para a avaliação automatizada das atividades. A ferramenta foi escolhida devido a sua ampla utilização atualmente tanto por professores e alunos. A versão gratuita foi utilizada considerando o critério de acessibilidade, a representatividade do uso real e com o objetivo de explorar as limitações intencionais do estudo. Foram utilizadas três formas de análise: Avaliação livre, avaliação com rubrica e avaliação com lista de verificação. Inicialmente as atividades foram avaliadas de forma livre, e em seguida foram avaliados com rubrica ou com lista de verificação

Avaliação Livre:

Neste tipo de avaliação a instrução para IA era avaliar as atividades atribuindo nota de 0 (zero) a 10 (dez). Os critérios utilizados para avaliação foram da própria IA sem a interferência dos docentes das disciplinas. A instrução dada ao Chat GPT para a análise livre foi: “Atribua nota de 1 a 10 pontos aos trabalhos. Justifique”.

Avaliação com Rubrica:

Na avaliação com rubrica, além dos arquivos dos relatórios, foi enviado à IA o documento contendo a rubrica correspondente. O arquivo incluía pelo menos oito critérios e cinco níveis, abordando aspectos como informações técnicas, formato, planejamento e questões reflexivas. O prompt utilizado foi: “Atribua uma nota de 1 a 10 aos trabalhos, utilizando a rubrica em anexo para a avaliação. Justifique sua atribuição”

Avaliação com lista de verificação:

Neste tipo de avaliação foi verificado se alguns itens pré-estabelecidos foram cumpridos pelo estudante. Cada item da lista correspondia a um aspecto que deveria estar presente no relatório, e a presença (ou ausência) desses itens orientava a atribuição da nota. Para isso foi anexado um documento contendo a lista de verificação das atividades e a IA foi instruída a realizar a avaliação considerando os critérios de lista de verificação e atribuir notas de 0 (zero) a 10 (dez). Não foi estabelecido pesos diferenciados pelo docente para os critérios de avaliação

2.4 Procedimentos de comparação

Para avaliar a correspondência entre as avaliações realizadas por docentes e por Inteligências Artificiais (IA), foram adotadas abordagens quantitativas e qualitativas. No que se refere à análise quantitativa, inicialmente foi organizada uma tabela com as notas atribuídas pelo professor, pela IA livre e pela IA com critérios. Essa tabela serviu de base para todas as análises subsequentes. Com esses dados, foram calculadas as médias das notas de cada categoria e o valor absoluto da diferença entre a avaliação feita pelo docente e pela IA (livre ou com critérios), a fim de obter um primeiro panorama sobre as discrepâncias entre as notas atribuídas.

Em seguida, com o objetivo de verificar a relação entre as avaliações, foi calculado o coeficiente de correlação de Pearson entre as notas do professor e da IA livre, e entre as notas do professor e da IA com critérios. Essa análise foi realizada utilizando o Jamovi (<https://www.jamovi.org/>), um software estatístico gratuito e de código aberto, que possibilitou calcular o grau de correlação e classificá-lo como muito forte, forte, moderada, fraca ou desprezível. Utilizando o mesmo software, foi aplicado o teste t pareado, com o objetivo de verificar se havia diferença estatisticamente significativa entre as médias das notas atribuídas pelos diferentes avaliadores. A interpretação dos valores de p seguiu os critérios

convencionais: valores de p inferiores a 0,05 indicam que a IA e o professor atribuem notas diferentes de forma consistente, enquanto valores iguais ou superiores a 0,05 indicam que a IA apresenta comportamento semelhante ao do professor na atribuição das notas.

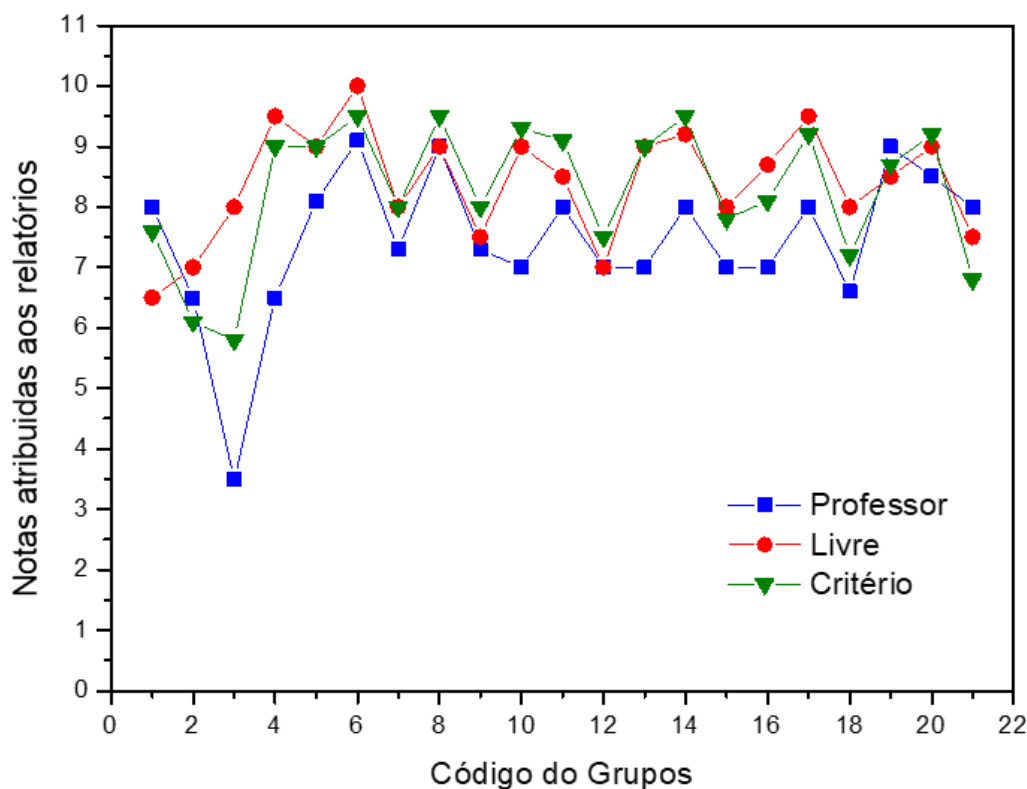
No que diz respeito à análise qualitativa, foram examinados os feedbacks fornecidos pelos professores e pelas IAs, em conjunto com as respectivas notas atribuídas. A análise buscou identificar o grau de convergência entre os comentários dos avaliadores, a coerência dos feedbacks gerados pelas IAs e a ocorrência de discrepâncias relevantes entre as observações feitas pelas IAs e pelos docentes, tanto na avaliação realizada pela IA em modo livre quanto com critérios. Essa abordagem permitiu avaliar não apenas as notas atribuídas, mas também a qualidade argumentativa e formativa do estudante.

3 RESULTADOS E DISCUSSÃO

3.1 Análise Quantitativa

A Figura 1 apresenta a distribuição das notas atribuídas pelo professor e pela IA (livre e com critérios) para os relatórios das disciplinas de Processamento de Termoplásticos e Processamento de Materiais Cerâmicos. O objetivo é comparar as diferenças entre avaliação feita por humanos e inteligência artificial e identificar padrões nas variações das notas. Para isso, as notas foram organizadas em faixas numéricas, representando os relatórios de cada grupo.

Figura 1. Distribuição das notas de relatórios de acordo com as disciplinas, (1-9) Processamento de termoplásticos e (10-21) Processamento de materiais cerâmicos para a avaliação feita pelo professor, IA livre e a IA com critérios



Fonte: Autoria própria.

Os resultados apresentados na Figura 1 revelam uma tendência: as notas atribuídas pela IA, tanto na avaliação livre quanto na avaliação com critérios, foram, em média,

superiores às concedidas pelo professor. Essa discrepância foi especialmente acentuada nos trabalhos que receberam avaliações docentes inferiores a 7 pontos, com diferenças que chegaram a 45%. Em contraste, para notas acima de 7, a variação entre as avaliações foi menos expressiva, oscilando entre 0% e 20%.

Ao analisar as médias gerais, constatou-se que a avaliação do professor resultou em uma nota média de 7,45 pontos, enquanto a IA atribuiu médias de 8,40 pontos (avaliação livre) e 8,28 pontos (avaliação com critério). Esses valores representam diferenças de 12% e 11%, respectivamente, em relação à avaliação docente. Analisando desse ponto de vista, a inclusão de critérios no prompt da IA não foi suficiente para alinhar significativamente suas avaliações aos critérios estabelecidos pelo professor, uma vez que a diferença entre os dois métodos de avaliação pela IA (livre e com critério) foi pouco significativa (apenas 1%).

Com base nesses resultados, surgem questões importantes sobre os critérios de rigor avaliativo adotados pela IA em comparação com os padrões docentes. A maior discrepância em notas mais baixas pode sugerir que a IA tende a ser mais benevolente na correção de trabalhos de desempenho insatisfatório, enquanto, em avaliações de melhor desempenho, sua concordância com as notas do professor é mais evidente.

Para avaliar o grau de concordância entre as notas do professor e da IA, calculou-se o coeficiente de correlação de Pearson entre os pares de avaliações (professor × IA livre; professor × IA com critérios). Os resultados revelaram diferenças significativas entre as avaliações realizadas pela inteligência artificial (IA) e as atribuídas pelos docentes. Na condição de avaliação livre, o coeficiente de correlação de Pearson entre as notas da IA e as dos professores foi de 0,30, indicando uma correlação fraca. Por outro lado, na condição orientada por critérios, o coeficiente aumentou para 0,67, sugerindo uma correlação moderada entre as avaliações.

O teste t pareado foi aplicado para verificar a significância estatística das diferenças entre as notas atribuídas pelo professor e pela IA levando em conta as avaliações dos mesmos trabalhos realizados pelos métodos citados. Apesar dessa melhora na correlação de Pearson quando a IA foi instruída com critérios previamente definidos pelos docentes, os resultados do teste t pareado demonstraram que as médias das notas atribuídas pelos professores e pelas IAs permanecem significativamente diferentes. Para a IA livre, o valor-p obtido foi de 0,003, enquanto para a IA orientada o valor-p foi inferior a 0,001. Esses valores indicam que, estatisticamente, as médias entre as avaliações do professor e da IA são significativamente diferentes.

Esses achados indicam que a orientação das ferramentas de IA com diretrizes claras melhora o alinhamento com os padrões docentes, mas não elimina as discrepâncias. A IA tende a seguir um padrão mais próximo ao do avaliador humano quando instruída, porém continua apresentando diferenças significativas em termos absolutos. Portanto, seu uso deve ser considerado como ferramenta de apoio e não como substituição da avaliação humana.

3.2 Análise Qualitativa

Foram analisadas as justificativas fornecidas pela IA para cada nota, possibilitando a avaliação da coerência e da adequação dos critérios utilizados. Além disso, essa abordagem permitiu identificar padrões na maneira como a IA estabelece as notas e interpreta os itens da rubrica e da lista de verificação, auxiliando na detecção de possíveis inconsistências ou tendências específicas no processo de avaliação.

Avaliação Livre

Na avaliação livre realizada pela IA, observou-se que os critérios estabelecidos estavam, em parte, alinhados com os traçados pelo professor, incluindo profundidade técnica, correção ortográfica, organização e estrutura do texto. No entanto, a IA penalizou os alunos pela ausência de alguns elementos não exigidos no trabalho, como dados financeiros e

protótipo, além de considerar os textos muito extensos, mesmo sendo relatórios técnicos que naturalmente exigem conteúdo detalhado para garantir compreensão adequada. Um aspecto positivo na avaliação feita pela IA foi a atribuição de notas e justificativas individualizadas, apresentando os resultados com a identificação do aluno, a nota recebida e uma descrição dos pontos fortes, seguida das áreas a melhorar e sugestões de aprimoramento. A avaliação utilizou parcialmente a técnica de Feedback Sanduíche (Hosseinpour *et al.*, 2024) que estrutura a devolutiva em três etapas: elogio inicial, crítica construtiva e incentivo final. Embora tenha contemplado apenas os dois primeiros aspectos, a IA poderia ser orientada a realizar a análise completa seguindo essa abordagem.

Avaliação com rubrica

Ao avaliar os trabalhos por meio de uma rubrica e comparar as avaliações feitas pelo professor e pela IA, critério por critério, algumas observações podem ser destacadas. A IA atribuiu no quesito informações técnicas, na maioria dos casos, uma nota mais alta do que o professor. Nos relatórios avaliados, os elementos técnicos exigidos foram desenvolvidos ao longo das aulas e eram fundamentais para a elaboração do texto escrito, mas como a IA utiliza uma rubrica genérica de avaliação, não identificou a ausência desses itens, o que resultou em notas mais baixas que foram atribuídas pelo professor. No item planejamento de ações, a IA concedeu uma nota inferior nesse critério, alegando falta de aprofundamento na viabilidade comercial do trabalho. No contexto deste curso de engenharia, as aulas priorizaram o entendimento do processo, sendo a parte comercial tratada apenas como um complemento, razão pela qual o professor frequentemente atribuiu notas mais altas nesse aspecto. A IA considerou o texto excessivamente técnico, apesar de se tratar de um trabalho acadêmico de engenharia, observação que provavelmente decorreu da ausência dessa especificação no prompt, influenciando na avaliação e resultando em uma nota menor atribuída pela IA. Embora o trabalho apresentasse poucos erros ortográficos, eles passaram despercebidos pela IA, mas foram identificados pelo professor, impactando na diferença da nota final. Alguns trabalhos não continham elementos essenciais presentes em relatórios, como conclusão e citações, ou não seguiam o formato adequado de relatório, e a IA não identificou a ausência desses itens, atribuindo notas entre 5 e 6 em tais casos, enquanto o professor atribui uma nota igual a zero.

Avaliação com Lista de Verificação

A análise das avaliações realizadas por meio de uma lista de verificação evidenciou comportamentos distintos da ferramenta de IA, mesmo quando os mesmos critérios foram utilizados em diferentes abordagens.

Inicialmente, ao ser fornecida apenas a lista de verificação como base, a IA reproduziu os critérios corretamente, estruturando sua análise em conformidade com os itens indicados pelo professor. Cada critério foi avaliado individualmente com uma nota de 0 a 10, e a nota final foi calculada como a média aritmética simples dessas avaliações. Esse modelo gerou uma estrutura de resposta compatível com a expectativa docente, embora a distribuição de notas por critério nem sempre refletisse a percepção do avaliador humano.

Em uma segunda configuração, utilizando o mesmo prompt acompanhado da lista de verificação, a IA atribuiu pesos diferenciados aos critérios, mesmo sem instrução explícita para tal. Nesse caso, a nota final passou a ser uma média ponderada, o que ocasionou distanciamentos maiores em relação às notas do professor. Um exemplo ilustrativo foi observado no critério “referências”: a IA atribuiu peso de apenas 5% a esse item, o que resultou em uma nota final de 9,3 para um aluno, enquanto a docente, que considerava esse critério mais relevante, atribuiu nota 7. Essa discrepância sugere que o peso subjetivo atribuído pelos avaliadores humanos a certos aspectos do trabalho não foi capturado adequadamente pela IA.

Por fim, uma terceira forma de apresentação dos resultados foi observada. Utilizando os mesmos critérios, a IA não atribuiu notas numéricas, mas utilizou indicadores visuais para

expressar o desempenho em cada item: sinal verde para itens plenamente atendidos, um símbolo de exclamação amarelo para critérios parcialmente atendidos e um X vermelho para os que não foram atendidos. Essa representação qualitativa, embora visualmente clara, dificulta a comparação direta com a avaliação docente, baseada em uma escala numérica. No entanto, pode ser útil como forma de feedback formativo.

Essas variações na forma de processamento e apresentação dos resultados demonstram que, mesmo com os mesmos insumos (lista de verificação e prompt), o comportamento da IA pode ser influenciado por interpretações internas do modelo. Isso reforça a necessidade de controle e verificação cuidadosa ao se empregar IAs em processos avaliativos.

4 CONSIDERAÇÕES FINAIS

Diante do objetivo de analisar a eficiência da utilização da inteligência artificial como ferramenta de apoio na correção de relatórios elaborados por estudantes dos cursos de Engenharia Metalúrgica e de Engenharia de Materiais da Universidade Federal Fluminense, as considerações finais apontam para importantes reflexões sobre a aplicabilidade, os benefícios e os limites dessa tecnologia no contexto educacional.

Entre os pontos positivos observados no uso da inteligência artificial na avaliação de relatórios, destaca-se a agilidade do processo, que leva apenas alguns segundos para ser concluído, a primeira análise, por exemplo, ocorreu em aproximadamente cinco segundos. Um aspecto relevante foi a capacidade da IA de oferecer *feedback* estruturado, destacando pontos fortes e aspectos a melhorar nas produções acadêmicas dos alunos. Esse tipo de retorno favorece uma revisão mais detalhada e ajustes direcionados, contribuindo significativamente para o aprimoramento da escrita e da organização dos trabalhos. Além disso, a IA sugeriu a geração de uma planilha com as notas e observações, facilitando o registro e o acompanhamento do desempenho discente. Os resultados apresentados incluíam nome e nota do aluno, comentários sobre os pontos positivos e os que demandavam melhorias, bem como sugestões coerentes e bem justificadas. A estrutura da proposta foi clara e articulada, e chamou atenção o fato de a IA mencionar se os trabalhos atendiam à terminologia técnica exigida. Como complemento, a ferramenta ainda ofereceu um resumo final das notas, sugerindo a geração de pareceres em PDF ou de comentários individuais para devolutiva aos alunos. A ficha de avaliação disponibilizada incluía a média geral, além dos destaques positivos e negativos e pontos de melhoria, mostrando-se uma alternativa promissora para auxiliar docentes na gestão de notas, acompanhamento de desempenho e preparação de feedbacks individualizados.

Por outro lado, a versão do ChatGPT utilizada neste trabalho apresentava algumas limitações, como a restrição de anexar arquivos apenas três vezes por dia, o que impactava o uso contínuo da ferramenta. Também foram observadas falhas na avaliação, como a ausência de penalização adequada para trabalhos que não continham elementos essenciais, como conclusão ou citações, ou que não seguiam o formato exigido de relatório, nesses casos, a IA não atribuiu nota zero, indicando que certos critérios podem não ter sido plenamente reconhecidos pelo sistema. Apesar de os textos conterem poucos erros ortográficos, nem todos foram identificados pela IA, o que comprometeu a precisão da avaliação. Além disso, houve registros de erros no arredondamento das notas, o que pode gerar inconsistências nas médias e diferenças injustas entre os alunos; por isso, recomenda-se verificar a coerência das notas atribuídas e incluir critérios claros na rubrica para mitigar esse tipo de falha. Outro ponto importante refere-se ao fato de a IA atribuir pesos próprios aos critérios avaliativos, que nem sempre refletem as prioridades do docente, uma rubrica bem detalhada e alinhada com os objetivos pedagógicos pode evitar distorções. Por fim, constatou-se uma limitação técnica do

sistema relacionada à impossibilidade de anexar documentos diretamente do Google Drive, exigindo que os arquivos estivessem previamente salvos no dispositivo local.

Conclui-se que a inteligência artificial apresenta potencial como ferramenta de apoio à correção de atividades acadêmicas, especialmente pela agilidade e pela qualidade do feedback gerado. No entanto, limitações como a atribuição automática de pesos, falhas na detecção de critérios e restrições técnicas indicam que seu uso ainda requer supervisão docente. Quando bem orientada, a IA pode complementar o trabalho do professor.

AGRADECIMENTOS

Agradecemos ao apoio institucional do Programa de Inovação e Assessoria Curricular (PROAIC) e do Departamento de Engenharia Metalúrgica e de Materiais (VMT) da Universidade Federal Fluminense, cuja colaboração foi essencial para o desenvolvimento e a realização deste trabalho. Estendemos também nosso agradecimento aos alunos que participaram ativamente das atividades propostas nas disciplinas envolvidas, contribuindo de forma significativa para os resultados alcançados.

REFERÊNCIAS

ALMEIDA, A. P.; LOPES, A. G.; ROCHA, M. C. M.; LINS, M. R. **Competências para educar na era da IA e os desafios do contexto brasileiro**. Jornal da USP, 13 ago. 2024. Disponível em: <https://jornal.usp.br/artigos/competencias-para-educar-na-era-da-ia-e-os-desafios-do-contexto-brasileiro/>. Acesso em: 29 de maio de 2025.

CASTELBLANCO, Gabriel; CRUZ-CASTRO, Laura; YANG, Zhenlin. **Performance of a Large-Language Model in scoring construction management capstone design projects**. Computer Applications in Engineering Education, v. 32, e22796, 2024.

CASTRO, Michele Marta Moraes; MACIEL, Cristiano. Historiográficos da Inteligência Artificial e suas Implicações na Educação. In: Anais do XXXII Seminário de Educação, 2024, Porto Alegre. **Anais**. Porto Alegre: SBC, 2024. p. 1273-1282. <https://doi.org/10.5753/semiedu.2024.32786>.

DU BOULAY, B. **Artificial intelligence in education and ethics**. In: ZAWACKI-RICHTER, O.; JUNG, I. (Eds.). Handbook of Open, Distance and Digital Education. Springer, 2023. p. 93-105. https://doi.org/10.1007/978-981-19-2080-6_6.

HOSSEINPOUR, A. *et al.* **Comparison of the effects of apprenticeship training by sandwich feedback and traditional methods on final-semester operating room technology students' perioperative competence and performance: a randomized, controlled trial**. BMC Medical Education, v. 24, n. 1, p. 578, 2024. <https://doi.org/10.1186/s12909-024-05598-6>.

JUKIEWICZ, M. **The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process**. Thinking Skills and Creativity, v. 52, p. 101522, 2024.

KEUNING, Hieke; HEEREN, Bastiaan; JEURING, Johan. **How Teachers Would Help Students to Improve Their Code**. In: INNOVATION AND TECHNOLOGY IN COMPUTER SCIENCE EDUCATION (ITiCSE '19), 2019, Aberdeen, Scotland UK. Proceedings. New York, NY, USA: ACM, 2019.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. **A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence**. 1955. Disponível em: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>. Acesso em: [data de acesso].

NGUYEN, A. *et al.* **Ethical principles for artificial intelligence in education**. Education and Information Technologies, v. 28, p. 4221–4241, 2022. <https://doi.org/10.1007/s10639-022-11316-w>.

PIMENTEL, Mariano; AZEVEDO, Viviane; CARVALHO, Felipe. **ChatGPT substituirá professoras e professores?** SBC Horizontes, 10 mar. 2023. ISSN 2175-92351. Disponível <https://horizontes.sbc.org.br/index.php/2023/03/chatgpt-substituira-professoras-e-professores/> Acesso em: 29 de maio de 2025.

UNESCO. **Guia para a IA generativa na educação e na pesquisa**. UNESCO, 2024.

EVALUATING THE PERFORMANCE OF ARTIFICIAL INTELLIGENCE IN THE ASSESSMENT OF ACADEMIC REPORTS IN ENGINEERING PROGRAM

Abstract: *Generative Artificial Intelligence (AI), particularly ChatGPT, has gained prominence in education through the automation of academic assessments. Despite its potential to speed up and standardize grading, it still faces limitations related to accuracy and bias. This study analyzes the performance of AI in grading academic reports in the Engineering Program at the Fluminense Federal University. Reports from courses on ceramic and polymeric materials processing were evaluated in three modalities: by professor, free-form AI, and AI with evaluation criteria. The AI-assigned grades were compared to those of professors using quantitative and qualitative analyses. Results indicate that although AI offers agility and structured feedback, significant discrepancies remain, especially in lower-performing assignments. It is concluded that, with proper guidance, AI can serve as a valuable support tool but does not replace human evaluation.*

Keywords: *artificial Intelligence, ChatGPT, automated assessment, engineering, report assessment*

