REDAMAX: Correção Automática de Redações do ENEM com Inteligência Artificial em Escolas Públicas

José Gustavo Araújo e Silva[1], Luciana Lopes Xavier[1], Josué Cordovil Medeiros[1], Marcio Aurelio dos Santos Alencar[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (IFAM)
Campus Coari-AM
Estrada Coari-Itapeuá, km 2 – Itamaraty – *Coari* – AM - Brasil

josegustavo101010@gmail.com,{luciana.xavier, josue.medeiros, marcio.alencar}@ifam.edu.br

Abstract. *The correction of essays in Brazil's National High School Exam (ENEM) remains a challenge in public schools, due to teacher overload and large class sizes. This paper presents REDAMAX, a web-based platform for automated essay scoring supported by large language models. The system evaluates essays according to the five official ENEM criteria, provides detailed feedback, and offers an optical character recognition (OCR) module for handwritten texts. This study validates the text-based correction pipeline using ten INEP 1000-score essays, employing the poolside/laguna-xs-2.1 model; the OCR pipeline was not exercised in this round and is reserved for future work. Results demonstrated consistent scoring and accurate identification of critical competencies such as intervention proposal and sociocultural repertoire, with an average processing time of 11.34 seconds and moderate token consumption per essay. The platform contributes to AI-assisted education and has replication potential in other public contexts.*

Resumo. *A correção de redações do Exame Nacional do Ensino Médio (ENEM) é um desafio recorrente em escolas públicas brasileiras, marcado por sobrecarga docente, turmas numerosas e escassez de feedback individualizado. Este artigo apresenta a plataforma REDAMAX, desenvolvida para avaliação automática de redações com apoio de modelos de linguagem de grande escala. O sistema atribui notas segundo os cinco critérios oficiais do exame, fornece feedback detalhado e disponibiliza um módulo de reconhecimento óptico de caracteres (OCR) para textos manuscritos. Neste estudo, validou-se especificamente o pipeline de correção textual, utilizando o modelo poolside/laguna-xs-2.1 sobre dez redações nota 1000 do INEP; o módulo de OCR não foi acionado nesta rodada experimental, permanecendo reservado para trabalhos futuros. Os resultados demonstraram consistência avaliativa e identificação precisa das competências mais críticas, como proposta de intervenção e repertório sociocultural, com tempo médio de processamento de 11,34 segundos por redação. Conclui-se que a solução representa contribuição tecnológica relevante para a educação pública, com potencial de replicação em outros contextos.*

1. Introdução

O Exame Nacional do Ensino Médio (ENEM) consolidou-se como o principal processo seletivo para ingresso no ensino superior brasileiro, e sua redação representa componente de peso significativo na nota final dos candidatos (INEP, 2025). A avaliação segue cinco

competências oficiais que exigem domínio da norma culta, compreensão do tema, organização argumentativa, coesão textual e proposta de intervenção (KLEIN; FONTANIVE, 2009).

No contexto das escolas públicas, professores enfrentam sobrecarga de trabalho e turmas numerosas, dificultando o acompanhamento individualizado (VIEGAS, 2022). Essa realidade compromete a preparação adequada dos alunos, que necessitam de prática sistemática e feedback contínuo para desenvolver habilidades complexas de escrita. A ausência de feedback formativo reforça desigualdades educacionais, já que estudantes de escolas privadas frequentemente recebem acompanhamento mais próximo (XAVIER, 2016). Nesse cenário, tecnologias de inteligência artificial surgem como alternativa viável para ampliar a oferta de correção e orientação pedagógica.

Avanços recentes em modelos de linguagem de grande escala possibilitam sistemas capazes de avaliar textos com sofisticação semântica e gramatical (BROWN et al., 2020). Esses modelos, quando integrados em plataformas educacionais, oferecem feedback imediato e personalizado, contribuindo para que o estudante compreenda seus erros e desenvolva estratégias de melhoria.

A plataforma REDAMAX foi concebida para atender às especificidades da redação do ENEM, integrando correção automática por múltiplos modelos de linguagem, reconhecimento óptico de caracteres e interface pedagógica acessível. O objetivo deste trabalho é investigar a viabilidade técnica e pedagógica de soluções de baixo custo operacional para escolas públicas. Este artigo descreve a arquitetura da plataforma, detalha os modelos de linguagem disponíveis, apresenta os resultados da validação do pipeline textual com redações nota 1000 do INEP — incluindo desempenho avaliativo e computacional — e discute limitações e perspectivas futuras, posicionando a solução como contribuição relevante para a educação assistida por inteligência artificial

2. Trabalhos Relacionados

A avaliação automática de redações tem sido objeto de estudo desde a década de 1960, com sistemas pioneiros como o Project Essay Grade (PAGE, 1966). Com o avanço do processamento de linguagem natural, surgiram soluções mais sofisticadas, como o Intelligent Essay Assessor (FOLTZ; LAHAM; LANDAUER, 1999), amplamente utilizado em contextos educacionais internacionais.

No Brasil, Fonseca et al. (2018) aplicaram redes neurais recorrentes e abordagens baseadas em features para avaliar competências linguísticas e argumentativas de redações do ENEM, obtendo correlação significativa com notas atribuídas por corretores humanos. Essa linha de pesquisa demonstrou a viabilidade de sistemas automáticos em contextos específicos. Já Marinho et al. (2022) propuseram uma abordagem de avaliação automática de redações estruturada diretamente em torno das competências do ENEM, alcançando melhorias nos índices de concordância com avaliadores humanos.

Os estudos recentes evidenciam o crescente emprego da inteligência artificial como suporte à produção e à avaliação de redações do ENEM. Ramos et al. (2026) realizaram uma

análise comparativa de diferentes ferramentas de IA generativa na correção de redações nota 1000, avaliando sua proximidade em relação às notas oficiais do exame. Já Reategui et al. (2025) desenvolveram o RIANA, um assistente baseado em Modelos de Linguagem de Grande Escala (LLMs) para apoiar a correção de redações e fornecer feedback aos estudantes. Por sua vez, Sachete et al. (2025) desenvolveram o EnemIA, ambiente que oferece devolutivas alinhadas às cinco competências do ENEM; a ferramenta foi aplicada experimentalmente em uma turma de ensino médio técnico e comparada a outra plataforma comercial de correção, evidenciando maior granularidade e especificidade no feedback gerado.

A plataforma REDAMAX diferencia-se por integrar correção automática multimodelo, reconhecimento óptico de caracteres e interface pedagógica voltada ao ensino médio público. Essa abordagem multifuncional representa contribuição original ao campo, com foco explícito nas especificidades do exame ENEM e no contexto das escolas públicas brasileiras.

3. Metodologia

A plataforma foi desenvolvida como aplicação web com módulos distintos para alunos, professores e administradores. Essa organização promove modularidade, facilita manutenção e estabelece limites claros de controle de acesso entre diferentes perfis de usuário.

O REDAMAX disponibiliza múltiplos Modelos de Linguagem de Grande Escala (LLMs) para a etapa de correção automática de redações, permitindo que a instituição selecione a alternativa mais adequada de acordo com critérios de desempenho, tempo de resposta e qualidade da avaliação. Os modelos de inteligência artificial integrados à plataforma e adotados neste projeto possuem acesso gratuito, não exigindo pagamento por uso, assinaturas ou licenciamento para sua utilização, o que reduz os custos de implantação e operação do sistema em instituições de ensino. Os modelos atualmente integrados são:

Correção textual: poolside/laguna-xs-2.1:free, nvidia/nemotron-nano-9b-v2:free, google/gemma-4-26b-a4b-it:free e nvidia/nemotron-3-super-120b-a12b:free;

Reconhecimento óptico de caracteres (OCR): nvidia/nemotron-nano-12b-v2-vl:free, modelo multimodal responsável pela conversão de redações manuscritas em texto digital.

O fluxo completo previsto para uso em sala de aula, quando a redação é produzida no caderno, inicia-se com a escrita manuscrita pelo estudante durante a aula; em seguida, a página é digitalizada pelo sistema e submetida ao modelo de OCR, que converte o texto manuscrito em texto digital; por fim, esse texto digital é processado por um dos modelos de correção, no mesmo pipeline aplicado às redações digitadas diretamente na plataforma. Essa arquitetura em duas etapas mantém a avaliação textual desacoplada da captura da escrita, favorecendo a manutenção e a substituição independente de cada modelo.

Na validação relatada neste artigo, empregou-se exclusivamente o pipeline de correção textual, com o modelo poolside/laguna-xs-2.1:free processando redações já digitadas, tal como extraídas das cartilhas do INEP. O módulo de OCR não foi acionado nesta etapa experimental, uma vez que o objetivo foi isolar a qualidade da avaliação textual em si, sem a variável adicional introduzida pela conversão de manuscritos.

O modelo de linguagem é orientado por um prompt estruturado, que estabelece explicitamente o papel do avaliador e os parâmetros de pontuação a serem seguidos:

"Você é um sistema especializado na avaliação de redações do Exame Nacional do Ensino Médio (ENEM). Avalie a redação fornecida com base nos cinco critérios abaixo. Para cada critério, forneça uma nota de 0 a 200. No final, calcule a nota total (soma das notas dos cinco critérios). Valorize repertórios produtivos, boa estrutura e argumentação consistente."

Em seguida, o prompt detalha as exigências oficiais de cada uma das cinco competências, conforme descritas na Cartilha do Participante da Redação do ENEM 2025 (INEP, 2025):

Competência I — domínio da modalidade escrita formal da língua portuguesa, avaliando adequação gramatical, ortográfica e sintática do texto;

Competência II — compreensão da proposta de redação e aplicação de conceitos de diferentes áreas do conhecimento no desenvolvimento do tema, a partir de repertório sociocultural produtivo, evitando cópia dos textos motivadores;

Competência III — seleção, organização, relação e interpretação de informações, fatos, opiniões e argumentos em defesa de um ponto de vista;

Competência IV — demonstração de conhecimento dos mecanismos linguísticos necessários à construção da argumentação, com uso adequado de recursos coesivos;

Competência V — elaboração de proposta de intervenção para o problema abordado, respeitando os direitos humanos e detalhando meio de execução, agente responsável, efeito esperado e, quando pertinente, detalhamento adicional.

A incorporação dessas exigências ao prompt busca reproduzir, na medida do possível, os parâmetros oficiais adotados pelos corretores humanos do exame, alinhando a saída do modelo à Matriz de Referência do INEP (INEP, 2025). Além da nota por competência, o sistema retorna um texto de feedback destacando os principais erros identificados em cada critério, orientando o estudante sobre pontos específicos de melhoria.

A interface de produção de redações da plataforma (Figura 1) destina-se ao processo de submissão e elaboração textual pelos estudantes. Nesta tela, o aluno pode selecionar o tema proposto, visualizar os textos motivadores e desenvolver sua redação diretamente no editor integrado. Além da digitação manual, a interface disponibiliza acesso à funcionalidade de OCR, permitindo a conversão futura de textos manuscritos em formato digital para análise. A

interface também oferece orientações relacionadas à estrutura da redação do ENEM, critérios avaliativos e acompanhamento do limite de linhas.



Figura 1 – Tela de produção de redação

O dashboard do professor (Figura 2) consolida as principais métricas da turma em um único painel, com indicadores-chave como total de alunos matriculados, total de redações corrigidas, média geral das notas e tempo médio de escrita, além de gráfico comparativo de desempenho por turma.

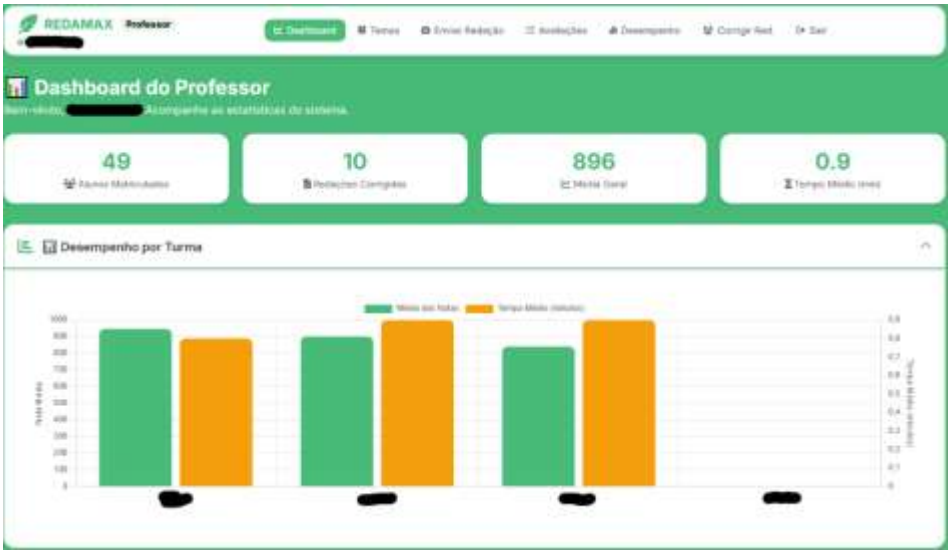


Figura 2 – Dashboard do Professor

A tela de gerenciamento de temas (Figura 3) permite ao professor criar, editar, ativar/desativar e excluir temas de redação, cada um composto por título, textos motivacionais, limite de linhas e status de ativação. Dashboard do Professor

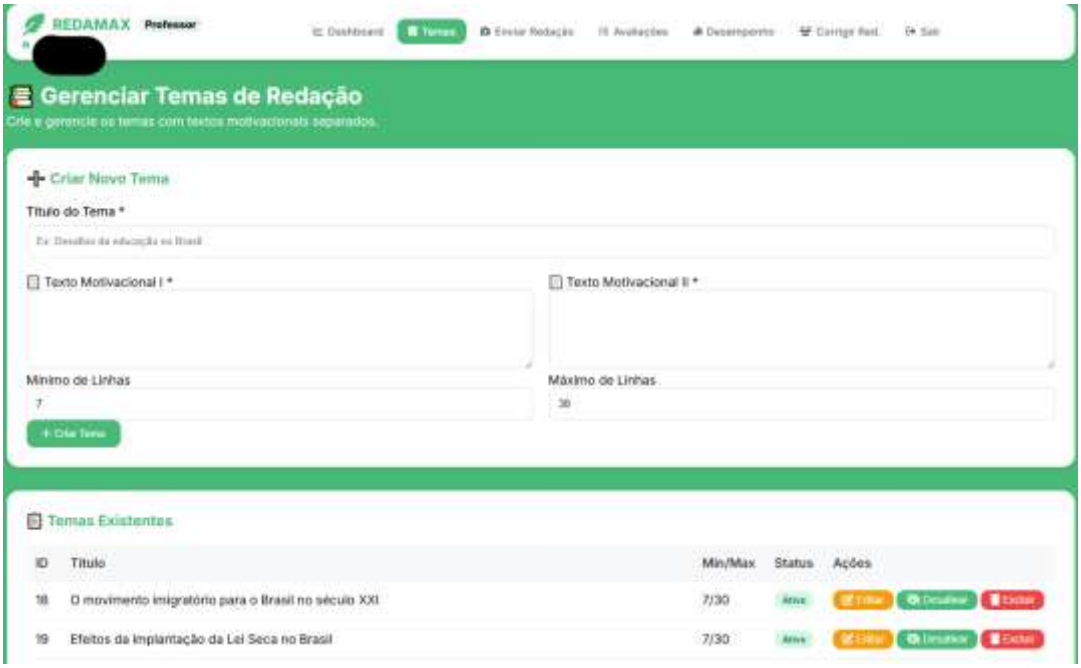


Figura 3 –Tela de gerenciamento de temas

A tela de avaliação de alunos (Figura 4) consolida o desempenho individual e coletivo dos estudantes, com lista de alunos, número de redações realizadas, média geral e média por competência (C1 a C5).

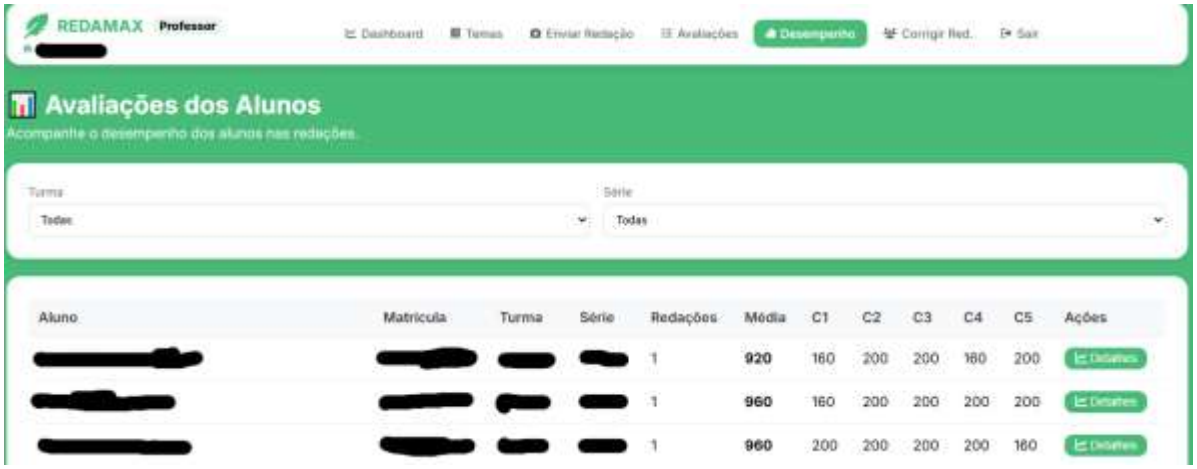


Figura 4 –Tela de desempenho dos alunos

Ao selecionar um aluno específico, o professor acessa um painel detalhado com gráfico de radar por competência (Figura 5) e histórico completo de redações, notas e feedback da IA.



Figura 5 – Painel detalhado das notas do aluno

4. Resultados

Para validar a eficácia do pipeline de correção textual, foram selecionadas dez redações (R1 a R10) com nota oficial 1000 publicadas pelo INEP, duas de cada edição do exame entre 2015 e 2019, abrangendo cinco temas distintos: A persistência da violência contra a mulher na sociedade brasileira (2015); Caminhos para combater a intolerância religiosa no Brasil (2016); Desafios para a formação educacional de surdos no Brasil (2017); Manipulação do comportamento do usuário pelo controle de dados na internet (2018); e Democratização do acesso ao cinema no Brasil (2019). A escolha de dois exemplares por ano e de temas variados teve como objetivo evitar que o resultado refletisse as particularidades de um único texto ou de um único tema, tornando a validação mais representativa.

As redações foram obtidas por meio das cartilhas oficiais do participante publicadas anualmente pelo INEP, que desde 2009 divulga exemplos de redações nota 1000 como referência pedagógica (INEP, 2025). Cada texto foi digitado integralmente e submetido ao pipeline textual do sistema — modelo poolside/laguna-xs-2.1:free — sem qualquer modificação de conteúdo, simulando o mesmo processo que um estudante realizaria ao usar a plataforma. O sistema processou cada redação de forma independente, sem acesso ao gabarito oficial, e retornou automaticamente as notas por competência (C1 a C5), a nota total e o feedback textual, com base exclusivamente no conteúdo submetido e nas rubricas codificadas no prompt.

Como critério de análise, adotou-se a comparação direta entre a nota atribuída pelo sistema e a nota oficial do INEP, calculando-se o erro da nota total para cada redação. A partir desses valores, foram calculadas a média das notas atribuídas pelo sistema e a média dos erros por competência e no total, permitindo identificar em quais critérios o sistema apresentou maior ou menor proximidade com o gabarito humano oficial.

Os resultados (Tabela 1) mostram que o REDAMAX atribuiu nota média de 896 pontos ao conjunto de redações cuja nota oficial é 1000. Seis das dez redações receberam nota igual ou superior a 920 pontos, indicando alta proximidade com o gabarito na maioria dos casos. Os dois maiores desvios ocorreram nas redações R6 (2017) e R8 (2018), ambas avaliadas com 760 pontos, com erro de 240 pontos em relação ao gabarito oficial.

Tabela 1 - Notas atribuídas pelo REDAMAX e comparação com o gabarito INEP

Redação	Ano	C1	C2	C3	C4	C5	Total	Erro
R1	2015	160	200	200	160	200	920	80
R2	2015	160	200	200	200	200	960	40
R3	2016	200	200	200	200	160	960	40
R4	2016	160	200	200	200	200	960	40
R5	2017	160	200	200	160	200	920	80
R6	2017	120	160	160	160	160	760	240
R7	2018	200	200	200	200	160	960	40
R8	2018	120	160	160	160	160	760	240
R9	2019	160	160	160	160	160	800	200
R10	2019	200	160	200	200	200	960	40
Média	—	164	184	188	180	180	896	104
Gabarito INEP	—	**200**	**200**	**200**	**200**	**200**	**1000**	—

A análise da média de erros por competência revela desempenho diferenciado entre os cinco critérios avaliados. A competência C3 (organização e progressão argumentativa) foi a que o sistema avaliou com maior precisão, com média de erro de 12 pontos, o que se explica pelo fato de a progressão argumentativa se manifestar de forma relativamente objetiva na estrutura do texto — presença de tese, desenvolvimento com argumentos encadeados e conclusão coerente são elementos que o modelo consegue identificar com consistência.

A competência C2 (compreensão do tema e uso de repertório sociocultural) apresentou média de erro de 16 pontos, segundo melhor resultado. O bom desempenho nessa competência indica que o sistema é capaz de identificar a pertinência temática do texto e reconhecer a presença de argumentos embasados em dados, fatos históricos ou referências culturais relevantes.

As competências C4 (coesão textual) e C5 (elaboração da proposta de intervenção) apresentaram média de erro idêntica de 20 pontos. No caso de C5, o resultado é especialmente relevante, pois a proposta de intervenção é o critério de maior objetividade estrutural no ENEM, exigindo a presença explícita de elementos como agente, ação, meio, efeito e detalhamento; o sistema demonstrou capacidade de verificar a presença ou ausência desses elementos com razoável precisão. Para C4, o resultado indica que o sistema identifica o uso de mecanismos coesivos, ainda que com margem de variação maior do que em C3.

A competência C1 (domínio da modalidade escrita formal da língua portuguesa) apresentou a maior média de erro, 36 pontos, com variação entre 0 e 80 pontos dependendo da redação. Esse resultado é explicável pela natureza dos textos submetidos: redações nota 1000 apresentam poucos ou nenhum desvio formal significativo, o que torna sutil a distinção entre uma C1 de 160 e uma de 200 pontos.

Além da qualidade avaliativa, mediu-se o desempenho computacional do modelo poolside/laguna-xs-2.1:free ao longo das dez correções realizadas. O tempo médio de processamento, desde o envio da redação até o retorno da nota por competência e do feedback textual completo, foi de 11,34 segundos por redação. Em termos de consumo de tokens, cada correção demandou, em média, 2.849 tokens de entrada, correspondentes ao prompt de sistema, aos critérios oficiais e ao texto da redação, e 1.608 tokens de saída, referentes às notas por competência e ao feedback textual.

Esses valores indicam que o custo computacional por correção é compatível com o uso em tempo real em sala de aula, como retorno assíncrono em plataformas de correção assistida ou como suporte a fluxos de revisão em lote, sem exigir infraestrutura de alto custo, característica desejável para adoção em escolas públicas com recursos limitados.

5. Discussão

Os resultados da validação indicam que o REDAMAX, operando com o modelo poolside/laguna-xs-2.1:free, apresenta comportamento consistente com o critério avaliativo do INEP para redações de alto desempenho. Além disso, a tendência de subestimação sistemática, observada pelo fato de o sistema não ter superestimado nenhuma das redações testadas, constitui uma característica pedagogicamente desejável em um sistema de feedback formativo, pois evita que o estudante receba avaliações infladas que não correspondam ao padrão real do exame.

Os desvios mais expressivos, observados nas redações R6 e R8, concentraram-se em temas de maior complexidade conceitual, relacionados à formação educacional de surdos e à manipulação de dados na internet. Nesses casos, o sistema avaliou com mais rigor competências como C1 e C2, possivelmente por interpretar de forma mais conservadora a adequação formal e o repertório sociocultural mobilizado, áreas em que o julgamento humano envolve maior subjetividade e contextualização cultural. Essa limitação aponta para a necessidade de ampliar o corpus de validação com redações de diferentes faixas de nota e temas variados.

Comparativamente à literatura, a média de erro total de 104 pontos em uma escala de 1000 (10,4%) situa o sistema em patamar compatível com sistemas AES (*Automated Essay Scoring*) de primeira geração descritos por Ramesh e Sanampudi (2022), que reportam erros médios entre 8% e 15% em avaliações holísticas. Sistemas mais recentes baseados em transformers pré-treinados alcançam erros entre 5% e 8%, indicando margem de melhoria viável para versões futuras do sistema, inclusive por meio da comparação com os demais modelos de maior capacidade já integrados à plataforma.

Do ponto de vista operacional, o tempo médio de 11,34 segundos e o consumo moderado de tokens por correção reforçam a viabilidade de uso do modelo poolside/laguna-xs-2.1:free em contextos de baixo orçamento, sem prejuízo relevante à qualidade avaliativa observada nesta validação.

6. Conclusões

Este artigo apresentou uma plataforma web para correção automatizada de redações do ENEM apoiada por inteligência artificial, desenvolvida e validada com redações nota 1000 do INEP. O sistema integra múltiplos modelos de linguagem de grande escala para correção textual, um modelo dedicado de reconhecimento óptico de caracteres para produções manuscritas e gerenciamento pedagógico, constituindo solução educacional completa para desenvolvimento da competência escrita de estudantes em preparação para o ENEM.

Os resultados demonstraram que o pipeline textual, operando com o modelo poolside/laguna-xs-2.1:free, é capaz de fornecer avaliações consistentes e feedback pedagogicamente relevante, com tempo médio de processamento de 11,34 segundos e consumo moderado de tokens por redação, a custo operacional reduzido para a instituição. A incorporação das exigências oficiais das cinco competências do ENEM ao prompt mostrou-se fundamental para aproximar o comportamento avaliativo do modelo daquele de um corretor humano experiente.

Como trabalhos futuros, pretende-se: (i) avaliar o pipeline completo de OCR, com redações efetivamente manuscritas por estudantes em sala de aula e digitalizadas pelo modelo nvidia/nemotron-nano-12b-v2-vl:free; (ii) comparar o desempenho avaliativo, o tempo de resposta e o consumo de tokens entre os quatro modelos textuais disponíveis na plataforma; e (iii) realizar avaliação controlada da concordância entre notas geradas pela plataforma e notas atribuídas por corretores humanos certificados, utilizando corpus de redações com gabarito oficial mais amplo. A solução apresenta potencial significativo de replicação em outros contextos educacionais públicos de nível médio.

Referências

BROWN, T. B. et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS), v. 33, p. 1877-1901, 2020.

FOLTZ, P. W.; LAHAM, D.; LANDAUER, T. K. The Intelligent Essay Assessor: Applications to Educational Technology. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, v. 1, n. 2, 1999.

FONSECA, E.; MEDEIROS, I.; KAMIKAWACHI, D.; BOKAN, A. Automatically Grading Brazilian Student Essays. In: INTERNATIONAL CONFERENCE ON THE COMPUTATIONAL PROCESSING OF PORTUGUESE (PROPOR), 13., 2018, Canela. Proceedings. Cham: Springer, 2018. p. 170-179.

INEP - INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. A Redação do Enem 2025: Cartilha do Participante. Brasília: MEC/INEP, 2025.

KLEIN, R.; FONTANIVE, N. Uma nova maneira de avaliar as competências escritoras na redação do ENEM. Ensaio: Avaliação e Políticas Públicas em Educação, v. 17, n. 65, p. 585-598, 2009.

MARINHO, J.; CORDEIRO, F.; ANCHIÊTA, R.; MOURA, R. Automated Essay Scoring: An Approach Based on ENEM Competencies. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC), 19., 2022, Campinas. Anais. Porto Alegre: SBC, 2022. p. 49-60.

PAGE, E. B. The Imminence of Grading Essays by Computer. Phi Delta Kappan, v. 47, n. 5, p. 238-243, 1966.

RAMESH, D.; SANAMPUDI, S. K. An Automated Essay Scoring Systems: A Systematic Literature Review. Artificial Intelligence Review, v. 55, p. 2495-2527, 2022.

RAMOS, J. L. C.; BRITO, H. H. B.; SILVA, J. C. S.; MARTINS, A. G.; RODRIGUES, R. L. Análise Comparativa de Ferramentas de IA Generativa na Correção de Redações Nota 1000 no ENEM. EaD em Foco, v. 16, n. 1, e2556, 2026.

REATEGUI, E. B. et al. RIANA: Desenvolvimento de assistente de correção para redação do ENEM com uso de Modelos de Linguagem de Grande Escala (LLM). In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (CBIE), 2025, Curitiba. Anais Estendidos. Porto Alegre: SBC, 2025. p. 248-251.

SACHETE, A. S.; LOIOLA, A. V. S. F.; PEREIRA, A. M.; ROSSI, F. D.; GOMES, R. S. EnemIA: correção de redações do Enem com Inteligência Artificial. Texto Livre, v. 18, e58426, 2025.

VIEGAS, M. F. Trabalhando todo o tempo: sobrecarga e intensificação no trabalho de professoras da educação básica. Educação e Pesquisa, v. 48, e205248, 2022.