

**CONFIABILIDADE DE FERRAMENTAS DE DETECÇÃO DE TEXTOS DE
INTELIGÊNCIA ARTIFICIAL PARA USO DOS DOCENTES DE FISIOTERAPIA
DA UFVJM EM CORREÇÃO DE AVALIAÇÕES E ATIVIDADES
ACADÊMICAS**

Ramon Mariano De Almeida (ramon.mariano@ufvjm.edu.br)

Ana Clara D'villa Gonçalves Barbosa (ana.dvilla@ufvjm.edu.br)

Jéssica Pinheiro Dos Santos (jessica.pinheiro@ufvjm.edu.br)

Douglas Novaes Bonifácio (douglas.novaes@ufvjm.edu.br)

Emanuel Rodrigues Pinheiro (emanuel.rodrigues@ufvjm.edu.br)

Murilo Xavier Oliveira (murilo.xavier@ufvjm.edu.br)

Marcus Alcantara (marcus.alcantara@ufvjm.edu.br)

Renato Guilherme Trede Filho (trede@ufvjm.edu.br)

O uso crescente de ferramentas de inteligência artificial (IA) na educação superior tem transformado a produção acadêmica, mas também gerado preocupações sobre a originalidade e a integridade das avaliações. Detectores de textos gerados por IA foram desenvolvidos como resposta a esse desafio, embora sua confiabilidade ainda seja questionada, especialmente em contextos não anglófonos. Este estudo teve como objetivo avaliar a confiabilidade e a acurácia de diferentes detectores de IA para aplicação docente na correção de textos acadêmicos no curso de Fisioterapia da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM). Foram

analisados seis textos: três redigidos por humanos e três gerados por modelos de IA (ChatGPT, DeepSeek e Gemini), avaliados por seis detectores (Copyleaks, AI Text Classifier, QuillBot AI, Originality, Scribbr e Plagiarismchecker). A análise estatística foi conduzida em linguagem Python (versão 3.12) utilizando as bibliotecas pandas, numpy, scikit-learn, scipy e matplotlib. O coeficiente de correlação intraclasse (ICC = 0,93) indicou excelente confiabilidade interinstrumental. As áreas sob as curvas ROC (AUC) variaram de 0,83 a 0,95, com destaque para o Originality (AUC = 0,95) e o AI Text Classifier (AUC = 0,94). Ao adotar limiar conservador de 95% de especificidade, ambas as ferramentas mantiveram sensibilidade de 0,67, reduzindo falsos positivos. Conclui-se que os detectores avaliados apresentam desempenho satisfatório, sobretudo o Originality e o AI Text Classifier, desde que utilizados de forma complementar ao julgamento docente, com limiares conservadores que minimizem o risco de erros éticos e interpretativos.

Palavras-chave: inteligência artificial; confiabilidade; detectores de texto; ética acadêmica; ensino superior.