INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO TRIÂNGULO MINEIRO
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
XV Seminário de Iniciação Científica e Inovação Tecnológica
Uberlândia-MG, 04 e 05 de novembro de 2025
ISSN 2316-9338

IAs Generativas na Programação: Um Estudo Comparativo da Qualidade do Código Sugerido por ChatGPT, Deep Seek e Gemini

Giovanna Fernandes[1]; Marcelo de Almeida Maia; Carlos Eduardo de Carvalho Dantas[2]

[1]Estudante, IFTM *Campus* Uberlândia Centro, MG, bolsista IFTM, giovannafernandes@estudante.iftm.edu.br
[2]Professor(a), IFTM *Campus* Uberlândia Centro, MG, Dr. Ciência da Computação, carloseduardodantas@iftm.edu.br

Trechos de código são pequenos fragmentos reutilizáveis que demonstram como resolver problemas específicos de programação, sendo amplamente buscados por desenvolvedores na web para apoiar tarefas como aprendizado, reutilização e desenvolvimento acelerado. Com o surgimento de Modelos de Linguagem de Grande Escala (LLMs), como ChatGPT, Gemini e DeepSeek, esses modelos passaram a ser utilizados não apenas para gerar novos trechos de código, mas também para apoiar atividades como refatoração e correção de bugs. Diante da concorrência entre diferentes LLMs, uma oportunidade de pesquisa consiste em avaliar sistematicamente a qualidade do código gerado, especialmente no que diz respeito à sua legibilidade, um aspecto fundamental na compreensão e manutenção de software. Este estudo avaliou a legibilidade de 981 trechos de código gerados por três LLMs, ChatGPT, DeepSeek e Gemini, em resposta a 327 consultas envolvendo tarefas de programação, utilizando as regras de legibilidade da ferramenta de análise estática SonarLint como critério. Também foi explorada uma abordagem preliminar que combina as recomendações do SonarLint com os LLMs para refatoração automática focada na melhoria da legibilidade. O ChatGPT apresentou um comportamento mais equilibrado, gerando trechos com menos avisos e corrigindo 76% dos casos afetados. O DeepSeek produziu mais trechos com problemas de legibilidade, mas teve o melhor desempenho na correção. Em contraste, o Gemini obteve o menor desempenho e introduziu mais novos avisos de legibilidade. Embora promissora, a integração entre LLMs e ferramentas de análise estática ainda enfrenta desafios, como a compreensão contextual das regras e a preservação de trechos relevantes de código. Este trabalho apresenta: (i) um estudo comparativo sobre a legibilidade dos trechos de código gerados pelos três LLMs; (ii) um catálogo com 54 regras do SonarLint utilizadas para avaliar melhorias de legibilidade; e (iii) uma análise sobre o quanto os LLMs são capazes de melhorar automaticamente a legibilidade do código, com base na correção, persistência ou introdução de novos avisos.

Palavras-chave: chatgpt. código-fonte. deepseek. gemini. inteligência artificial. programação.

Apoio: IFTM.