

**XXI CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE FEDERAL DE CAMPINA
GRANDE**



***INVESTIGAÇÃO DO USO DE MODELOS DE LINGUAGEM DE GRANDE ESCALA
NA PESQUISA EM ENGENHARIA DE SOFTWARE***

Carolina Martins Freitas¹, João Arthur Brunet Monteiro²

RESUMO

Os Modelos de Linguagem de Grande Escala (LLMs) têm impactado de forma expressiva diversos domínios, incluindo a Engenharia de Software. Com sua capacidade de processar e gerar grandes volumes de dados textuais, esses modelos vêm sendo usados para automatizar processos que, até então, dependiam fortemente da intervenção humana, como a revisão de código. No entanto, apesar do avanço nas pesquisas, ainda há uma lacuna no entendimento de como esses modelos estão sendo aplicados na revisão de código. Esta pesquisa se propõe a preencher essa lacuna, investigando e caracterizando o uso de LLMs na automação da revisão de código por meio de uma revisão sistemática da literatura. Foram selecionados e analisados 25 artigos, provenientes de diferentes fontes, para responder a cinco questões de pesquisa. Na primeira questão de pesquisa, foram apresentados e classificados os principais LLMs utilizados nos estudos analisados. Na segunda questão, foram examinadas as subtarefas do processo de revisão de código nas quais esses modelos têm sido aplicados. A terceira questão explorou as estratégias de otimização adotadas para melhorar o desempenho dos modelos, como fine-tuning e prompt engineering. Na quarta questão, investigou-se o impacto de diferentes representações de input, como Persona Prompt Patterns e Marcadores Diff, no desempenho dos LLMs. Por fim, a quinta questão discutiu e categorizou as principais métricas de avaliação de desempenho dos modelos, com destaque para as métricas CodeBLEU, BLEU e F1-score. A partir das respostas a essas questões, foi possível traçar um panorama do estado da arte atual do uso de LLMs na automação da revisão de código, oferecendo uma referência para futuras investigações sobre o tema.

Palavras-chave: Aprendizagem de máquina, LLMs, Modelos de Linguagem de Grande Escala, Processamento de Linguagem Natural, Revisão de Código, Engenharia de Software.

¹Aluna do curso de Ciência da Computação, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail:carolina.freitas@ccc.ufcg.edu.br

²Doutor, professor, Departamento de Sistemas e Computação, UFCG, Campina Grande, PB, e-mail: joao.arthur@computacao.ufcg.edu.br