DAFTAR PUSTAKA

[1] H. M. Abushama, H. A. Alassam, and F. A. Elhaj, “The effect of Test-Driven Development and Behavior-Driven Development on Project Success Factors: A Systemantic Literature Review Based Study,” 2020.

[2] K. Karlsson and J. Gunnarsson, “Test-Driven Development Using LLM,” 2024.

[3] N. Pombo and C. Martins, “Test driven development in action: Case study of a cross-platform web application,” *EUROCON 2021 - 19th IEEE Int. Conf. Smart Technol. Proc.*, no. July, pp. 352–356, 2021, doi: 10.1109/EUROCON52738.2021.9535554.

[4] D. Staegemann *et al.*, “A Literature Review on the Challenges of Applying Test-Driven Development in Software Engineering,” *Complex Syst. Informatics Model. Q.*, vol. 2022, no. 31, pp. 18–28, 2022, doi: 10.7250/csimq.2022-31.02.

[5] Z. Liu, Y. Tang, X. Luo, Y. Zhou, and L. F. Zhang, “No Need to Lift a Finger Anymore? Assessing the Quality of Code Generation by ChatGPT,” *IEEE Trans. Softw. Eng.*, vol. 50, no. 6, pp. 1548–1584, 2024, doi: 10.1109/TSE.2024.3392499.

[6] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, “A Survey on Large Language Model (LLM) Security and Privacy: The Good, The Bad, and The Ugly,” *High-Confidence Comput.*, vol. 4, no. 2, p. 100211, 2024, doi: 10.1016/j.hcc.2024.100211.

[7] L. Chen, “Research on Code Generation Technology based on LLM Pre - training,” vol. 10, no. 1, 2024.

[8] A. Grattafiori *et al.*, “The Llama 3 Herd of Models,” pp. 1–92, 2024, [Online]. Available: http://arxiv.org/abs/2407.21783.

[9] M. M. Moe and K. K. Oo, “Evaluation of Quality, Productivity, and

Defect by applying Test-Driven Development to perform Unit Tests," *2020 IEEE 9th Glob. Conf. Consum. Electron. GCCE 2020*, pp. 435–436, 2020, doi: 10.1109/GCCE50665.2020.9291950.

[10] Z. Xue *et al.*, "LLM4Fin: Fully Automating LLM-Powered Test Case Generation for FinTech Software Acceptance Testing," *ISSTA 2024 - Proc. 33rd ACM SIGSOFT Int. Symp. Softw. Test. Anal.*, pp. 1643–1655, 2024, doi: 10.1145/3650212.3680388.

[11] S. Kang, J. Yoon, N. Askarbekkyzy, and S. Yoo, "Evaluating Diverse Large Language Models for Automatic and General Bug Reproduction," *IEEE Trans. Softw. Eng.*, vol. 50, no. 10, pp. 2677–2694, 2024, doi: 10.1109/TSE.2024.3450837.

[12] M. Schafer, S. Nadi, A. Eghbali, and F. Tip, "An Empirical Evaluation of Using Large Language Models for Automated Unit Test Generation," *IEEE Trans. Softw. Eng.*, vol. 50, no. 1, pp. 85–105, 2024, doi: 10.1109/TSE.2023.3334955.

[13] S. Piya and A. Sullivan, "LLM4TDD: Best Practices for Test Driven Development Using Large Language Models," *2024 IEEE/ACM Int. Work. Large Lang. Model. Code*, pp. 14–21, 2024, doi: 10.1145/3643795.3648382.

[14] A. M. Suruj, "A Compact Guide to Large Language Models," no. November, 2023.

[15] J. Yang *et al.*, "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 6, pp. 1–24, 2024, doi: 10.1145/3649506.

[16] H. Naveed *et al.*, "A Comprehensive Overview of Large Language Models (LLMs)," *Int. J. Multidiscip. Res.*, vol. 7, no. 1, 2024, doi: 10.36948/ijfmr.2025.v07i01.34609.

[17] R. Pan, M. Kim, R. Krishna, R. Pavuluri, and S. Sinha, "Multi-language

Unit Test Generation using LLMs," 2024, [Online]. Available: http://arxiv.org/abs/2409.03093.

[18] P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications," 2024, [Online]. Available: http://arxiv.org/abs/2402.07927.

[19] J. Xiang *et al.*, "Self-Supervised Prompt Optimization," 2025, [Online]. Available: http://arxiv.org/abs/2502.06855.

[20] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," *Proc. 40th Annu. Meet. Assoc. Comput. Linguist. (ACL),* pp. 311–318, 2002.

[21] S. Ren *et al.*, "CodeBLEU: a Method for Automatic Evaluation of Code Synthesis," vol. 1949, no. Weaver 1955, 2020, [Online]. Available: http://arxiv.org/abs/2009.10297.

[22] M. Popović, "Chrf: Character n-gram f-score for automatic mt evaluation," *10th Work. Stat. Mach. Transl. WMT 2015 2015 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2015 - Proc.*, no. September, pp. 392–395, 2015, doi: 10.18653/v1/w15-3049.

[23] S. Lu *et al.*, "CodeXGLUE: A Machine Learning Benchmark Dataset for Code Understanding and Generation," *Adv. Neural Inf. Process. Syst.*, 2021.