Master Thesis Topic
Automated Evolution of
Software Engineering agents

Motivation and Background:
Software Engineering (SE) agents/bots [1, 2] promise great practical support for software developers, e.g., by automatically fixing reported issues or automated feature additions. However, similar to human software developers, these systems can make mistakes. Human developers can learn from mistakes, but how does it work for SE bots? This thesis project explores the evolution of SE bots and how it can be supported by automated techniques.

Student Task and Responsibilities:
- Make yourself familiar with the state-of-the-art LLM-based APR techniques and the current LLM-based SE agents.
- Explore the concept of self-evolution of LLMs [3] in the context of SE bots, and systematically explore how SE bots can learn from mistakes.
- Focus first on benchmarks like Defects4J benchmark (https://github.com/rjust/defects4j), and then also focus on more sophisticated SE tasks, e.g., by investigating the issues in SWE-bench (https://www.swebench.com/).
- Design/select evaluation metrics and conduct a thorough evaluation of your approach.
- Analyze the results and document your findings.

Deliverables:
- Concept and prototypical implementation for the automated evolution of SE bots
- Evaluation artifacts (dataset, tools, etc.)
- Documented findings of the conducted experiments

Pre-Requisites: (Programming Languages, OS, Skills, Papers, etc)
A strong background in LLMs, Java, and Python is helpful for this project.

[1] I. Bouzenia, P. Devanbu, and M. Pradel, "RepairAgent: An Autonomous, LLM-Based Agent for Program Repair," Mar. 25, 2024. http://arxiv.org/abs/2403.17134

[2] Y. Zhang, H. Ruan, Z. Fan, and A. Roychoudhury. "AutoCodeRover: Autonomous Program Improvement". In Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2024). ACM, NY, USA, 1592–1604. https://doi.org/10.1145/3650212.3680384

[3] Tao, Z., Lin, T.E., Chen, X., Li, H., Wu, Y., Li, Y., Jin, Z., Huang, F., Tao, D. and Zhou, J., 2024. A survey on self-evolution of large language models. https://arxiv.org/pdf/2404.14387

Contacts
Prof. Dr. Yannic Noller (`sq-office@rub.de`)
Software Quality group, Faculty of Computer Science, Ruhr University of Bochum