Research Statement

I aim to develop AI-driven, developer-focused solutions that streamline the production of high-quality and secure software systems.

Research Experience

I am keen to develop practical solutions to real-world software engineering challenges. In particular, I use empirical methods, program analysis, and AI techniques to develop tools that help developers implement secure software engineering practices. My work has been deeply collaborative, leading to strong relationships with researchers and industry leaders around the world. This has resulted in more than 80 peer-reviewed papers, with an h-index of 23, and more than half published in well-respected, A-ranked international venues. Notably, I have collaborated with leading organisations such as TeamViewer, Forescout, Siemens, and Google. A key highlight was our partnership with Google to incorporate advanced security checks into Android Studio, which culminated in a successful SNSF project funded with CHF 368,517. This collaboration has had a tangible impact on both the academic community and the software industry.

Research Plan

The use of Large Language Models (LLMs) such as ChatGPT has become increasingly popular across disciplines. In software development, models like GitHub Copilot and Code Llama are widely adopted, driving the creation of larger and more complex systems, while accelerating their evolution at an unprecedented pace.

Studies show that developers often place excessive trust in these models, even though hallucinations and non-determinism create major barriers to reliable use. These challenges become even more pronounced in software security, where expert knowledge and trustworthy data are limited. Recent work further shows that LLMs behave far less consistently on security-critical tasks in this domain, and their effectiveness in real-world environments frequently falls short of expectations.

Over the next five years, I aim to establish a research group that will rigorously evaluate the strengths and weaknesses of large language models for engineering high-quality and secure software systems. We will investigate how these models perform in practice and develop scalable tools that can effectively assist developers in detecting and patching vulnerabilities within today's fast-paced and increasingly complex development landscape. In particular, we will focus on the following key objectives:

- Benchmark Development: We will create unbiased, representative benchmarks based on real-world scenarios that LLMs have not been exposed to, ensuring these benchmarks mirror the true diversity and complexity of software engineering tasks.
- Assessment: We will study both proprietary and open-source LLMs to evaluate their effectiveness in tasks such as secure code generation, vulnerability detection, and patching. We will examine how performance varies across different development contexts, considering factors such as environment and developer expertise.
- Model Improvement: Building on our evaluation insights, we will fine-tune or develop models tailored to secure software engineering. Our approach will focus on reducing errors and hallucinations through constraint-based techniques, enhancing determinism via structured prompting, and improving reliability using task-specific evaluation benchmarks.
- Tool Integration: We will leverage the improved models and insights from our evaluation to develop intelligent tools that provide real-time support for tasks such as coding, program analysis, and code review. Where appropriate, the models will be combined with traditional program analysis tools to create robust, hybrid solutions. We will rigorously evaluate our tools through controlled experiments and real-world case studies, comparing their effectiveness against current best practices and ensuring their safe integration into real-world development workflows.

This agenda sits at the intersection of program analysis, AI, security, and human–computer interaction, creating substantial opportunities for interdisciplinary collaboration and high-impact results. In the long term, I envision the creation of a holistic intelligent system that supports the entire software development lifecycle. It will continuously monitor development sessions, offer tailored advice on programming, debugging, and testing tasks, and streamline the creation of high-quality and secure software systems.

In the pursuit of this research, I will continue to cultivate and expand collaborations with academic and industry leaders. I will actively contribute to national and international research projects, disseminate our findings through top-tier journals and conferences, and work alongside industry partners to ensure our innovations are adopted in real-world settings.

I am confident that my research will make significant academic contributions and societal impact, advancing the development of high-quality and secure software systems.