

How to fine tune open-source Large Language Models (LLMs) with companies own software development data for personalised use cases

Presenter
Md Mahade Hasan¹
mdmahade.hasan@tuni.fi

Supervised by,
Prof. Pekka Abrahamsson¹
pekka.abrahamsson@tuni.fi

1. Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Finland

Research Questions:

- **RQ1:** What are the most prominent open-source LLMs for Software Engineering(SE), and how do they compare to proprietary models in terms of capabilities and limitations?
- **RQ2:** What are the most effective techniques to fine-tune open-source LLMs with private data to optimize their performance for specific software engineering use cases?
- **RQ3:** What are the key challenges of using fine-tuned LLMs in software engineering, and what strategies can help address issues like model accuracy, biases, and scalability?

The comic illustrates a dialogue between a user and AI models. The user asks how to build an AI tool for software testing that understands company practices. The AI models respond that they can help but need to be trained with company data. The user asks how to do it, and the AI models explain that they need to process company data on their servers. The user expresses concern about data privacy and not trusting anyone. The AI models reassure the user that they are open-source LLM models and can be used in a controlled environment to ensure data privacy. The user asks for a possible solution, and the AI models suggest that they are fine-tuned with the user's documents and ready to help. The user asks for a test case for the ABC feature of the XYZ project.