

CS 390 and 490 Proposals

1 Enhancing Software Analytics with Fine-Tuned LLMs and Prompt Engineering

Large Language Models (LLMs), like GPT-4 and OpenAI Codex, have shown remarkable abilities in natural language processing, code generation, and various software engineering tasks. In this project, we aim to explore how LLMs can be fine-tuned to deliver actionable insights into software analytics, particularly for code reviews, bug tracking, and developer productivity.

Students will focus on fine-tuning an LLM model with software engineering datasets, applying advanced prompt engineering techniques to create a software analytics tool. The tool will generate meaningful metrics and reports, automate issue classification, and provide recommendations for improving software development processes. The project will allow students to experiment with model fine-tuning, prompt engineering, and integrating LLM-powered insights into a practical software analytics platform.

2 Analyzing the Effectiveness of GitHub CoPilot

GitHub Copilot draws context from comments and code, and suggests individual lines and whole functions instantly. GitHub Copilot is powered by OpenAI Codex, a new AI system created by OpenAI. We would like to explore the limits of GitHub CoPilot by conducting several experiments.

3 Data science for Software Engineering

Software analytics aims to obtain insightful and actionable information from software artifacts that help practitioners accomplish tasks related to software development, systems, and users.” [1] With software analytics, the main idea is mining the big data in multiple software repositories to produce insights for decision makers in the software development process. The project has the following stages;

Getting familiar with the topic: This requires some guided reading about Software Analytics and Data science.

Research: First step is finding the proper software analytics question (s) that have practical impact to software practitioners. Then, you will investigate /combine various data mining techniques (machine learning algorithms, deep learning, big data graph visualization etc.) to answer the software analytics question.

Implementation: Application of the right data science technique to the right data science question

Experimentation: You will evaluate your approach based on available open source repositories.

[1] D. Zhang, S. han, Y. Dan, J.-G. Lou, H Zhang: ”Software Analytics in Practice”. IEEE Software, Sept./Oct. 2013, pp. 30-35.

Please check <https://www.cs.bilkent.edu.tr/cs490/current/index.html> and <https://www.cs.bilkent.edu.tr/cs390/current/> for eligibility. If you are interested, please send me an email with your CV and transcript (eraytuzun@cs.bilkent.edu.tr)