Assessing Large Language Models Trained on Source Code Using the LASSO Platform

Team Project

Chair of Software Engineering / Marcus Kessel

Fall Semester 2023
LLMs (and Chatbots) for Source Code

Trained on massive amounts of open source code

- OpenAI
- ChatGPT
- Codex
- GitHub Copilot
- tabnine
- StarCoder

... many more
Code & Test Generation Tasks

Code Generation (Program Synthesis)

GitHub Copilot

Test Generation (Inputs and Outputs)
Jack of all Trades, Master of None (?)

“… the robots are coming …”

“… they replace developers …”

“… hallucinating bullsh** …”

“… frees from boring tasks …”

/r/ProgrammerHumor
Assessing Code LLMs

State-of-the-Art Benchmarks
Curated Coding Task Collections → HumanEval, MBPP, MultiPL-E etc.

Measure and Analyze Functional/Non-Functional Behaviour

Delivers expected functionality?
High-quality code?
Better than Code Search?
Code Attribution?
Risks?
Goal (1)

- LASSO is a leading edge software observatorium that allows advanced search and analysis techniques to be applied to “big code”. Among other things, this simplifies experimentation and the validation of tools and software engineering approaches.

- The goal of this team project is to study leading-edge code LLMs (focus: code and test generation tasks) with the help of the LASSO platform. This includes –
  - Integrating state-of-the-art Code LLMs into LASSO to enable comparisons
  - Integrating/running established benchmarks (coding tasks)
  - Setting up analysis pipelines in LASSO’s scripting language, LSL, to automate the experimentation process including –
    - Test-Driven Assessment of Functional Behaviour
    - Code Quality Measurements
    - Code Attribution (code shared with original data)
    - Comparison with Code Search / Recommendation
Goal (2)

- **Participants**
  - 6 students

- **Length**
  - 6 months

- **Prerequisites**
  - (Java) Programming
  - Fundamental understanding in machine learning

- **Language**
  - English

- **Organisation**
  - Goals and timetable defined by agreement with the supervisor

- **Applicable to MMDS**: yes

- **Online**: By agreement

- **Supervisor**
  - Marcus Kessel