

Velké jazykové modely v softwarovém inženýrství: human-in-the-loop rámec pro kvalitu, bezpečnost a nasazení

Školitel: doc. Ing- Radek Šilhavý, Ph.D.

Konzultant: doc. Ing- Petr Šilhavý, Ph.D.

Ústav fakulty: Ústav počítačových a komunikačních systémů

Studijní program: Informační technologie

Anotace:

Velké jazykové modely (LLM) rychle pronikají do softwarového inženýrství a mění způsob, jakým vývojové týmy pracují s požadavky, kódem, testy, dokumentací i údržbou. Současně však přinášejí specifická rizika – zejména nekonzistenci výstupů, halucinace, obtížnou auditovatelnost, rizika úniku citlivých informací a nové bezpečnostní slabiny. Nedílnou součástí odpovědného nasazení se proto stává přístup human-in-the-loop, ve kterém člověk cíleně řídí, kontroluje a validuje výstupy modelu a kde jsou jasně definovány kontrolní body, role a metriky kvality.

Cílem disertační práce je (i) systematicky zmapovat a analyzovat využití LLM napříč životním cyklem vývoje software a identifikovat oblasti s nejvyšším potenciálem a současně s nejvyššími riziky, (ii) navrhnout metodický human-in-the-loop rámec pro integraci LLM do softwarového inženýrství (workflow, kontrolní mechanismy, validační postupy, metriky a auditní stopa) a (iii) navržený rámec empiricky ověřit na vybraných úlohách (např. generování/úpravy kódu, opravy chyb, návrh testů, analýza incidentů či práce s požadavky) s využitím reprodukovatelných benchmarků a reálných artefaktů. Pro evaluaci schopností LLM v realistických SE úlohách lze využít i existující standardizované benchmarky typu SWE-bench.

Hlavní oblasti výzkumu zahrnují:

- Systematickou analýzu současného stavu LLM v softwarovém inženýrství a syntézu typických úloh, přínosů a limitů (včetně „LLM4SE“ přehledů a roadmap).
- Návrh human-in-the-loop metodiky: role člověka v rozhodovacích bodech, strategie verifikace (testy, statická analýza, pravidla), dokumentace a auditovatelnost.
- Návrh praktických technik integrace: šablony interakce s LLM (prompting patterns), řízení kontextu, práce s nejistotou a řízení kvality výstupů v SE pipeline.
- Experimentální ověření: srovnání proti baseline bez LLM / bez HITL, přínos na kvalitě (např. correctness), produktivitě, robustnosti a bezpečnosti, včetně reprodukovatelnosti na benchmarkových úlohách a reálných datech.

Literatura:

- [1] SOMMERVILLE, Ian. Softwarové inženýrství. Brno: Computer Press, 2013, 680 s. ISBN 9788025138267.
- [2] Barenkamp, M., Rebstadt, J. & Thomas, O. Applications of AI in classical software engineering. AI Perspect 2, 1 (2020). <https://doi.org/10.1186/s42467-020-00005-4>
- [3] MAKRIDAKIS, Spyros. The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures, 2017, 90: 46-60.

[4] M. Harman, "The role of Artificial Intelligence in Software Engineering," 2012 First International Workshop on Realizing AI Synergies in Software Engineering (RAISE), Zurich, Switzerland, 2012, pp. 1-6, doi: 10.1109/RAISE.2012.6227961.

[5] FELDT, Robert; DE OLIVEIRA NETO, Francisco Gomes; TORKAR, Richard. Ways of applying artificial intelligence in software engineering. In: 2018 IEEE/ACM 6th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE). IEEE, 2018. p. 35-41.