

# Jiawei Li

jiawl28@uci.edu

Jiawei Li has extensive research experience in utilizing **Machine Learning**, **Deep Learning**, and **Large Language Models** to improve software quality. He successfully adapted **ReAct prompting** technique on GPT-4 to automatically generate commit messages by enabling it to engage in reasoning while interacting with the software repository through actions. The quality of the generated messages is considered to be better than humans and state-of-the-art techniques. The work has won **ACM SIGSOFT Distinguished Paper Award**. In addition, he built an **ensemble Machine Learning model (Bi-LSTM, Bi-GRU, and XGBoost)** to evaluate commit message quality, outperforming other existing models. Moreover, he has experience applying **static program analysis** techniques to detect software quality issues. Also, he has applied various empirical analysis techniques in his publications, including **Card Sorting**, **Open Coding**, and **Survey**.

EDUCATION	<b>Ph.D. Software Engineering</b> , University of California, Irvine (exp.) 2026
	<b>M.S. Computer Science</b> , University of California, Irvine 2020
	<b>B.Eng. Software Engineering</b> , Xi'an Jiaotong University 2019
SERVICES	<b>Reviewer (Full length technical paper)</b> , Automated Software Engineering 2025
	<b>Reviewer</b> , 2025 15th Berlin Machine Learning Workshop 2025
	<b>Shadow Research Track Program Committee Member</b> , 2026 48th IEEE/ACM International Conference on Software Engineering (ICSE) 2025
	<b>Reviewer (Full length technical paper)</b> , IEEE Transactions on Reliability 2025
	<b>Shadow Research Track Program Committee Member</b> , 2025 47th IEEE/ACM International Conference on Software Engineering (ICSE) 2024
	<b>Program Committee Member (Artifact Evaluation)</b> , 2025 47th IEEE/ACM International Conference on Software Engineering (ICSE) 2024
	<b>Reviewer (Full length technical paper)</b> , Transactions on Software Engineering and Methodology (TO SEM) 2024
	<b>Graduate Mentor</b> , Global Research Experience in Artificial Intelligence (GREAT) Program 2023
	<b>Graduate Mentor</b> , Global Research Experience in Artificial Intelligence (GREAT) Program 2022
WORK	<b>Applied Scientist Intern</b> , Amazon Inc. 2025
EXPERIENCE	<b>Teaching Assistant/Graduate Student Researcher</b> , University of California, Irvine 2021-Present
SELECTED PROJECTS	<b>Commit Message Generation Leveraging Reasoning and Action of Large Language Model</b> 2024 <ul style="list-style-type: none"><li>The published paper has won <b>ACM SIGSOFT Distinguished Paper Award</b>.</li><li>Built a <b>GPT-4-powered</b> commit message generation approach using <b>LangChain</b> that has access to a diverse and large software context associated with the code change.</li><li>Enabled GPT-4 to engage in reasoning while interacting with the external world through various actions using <b>ReAct prompting</b>. GPT-4 serves as a reasoning engine, determining the actions to take and their sequence. Actions can involve using a “tool” which represents a function for collecting external information to support intermediate reasoning and generating the commit message.</li><li>Achieved an average improvement of <b>30.2%</b> over human-written messages and a <b>71.6%</b> improvement over state-of-the-art commit message generation methods.</li></ul> <b>Analysis of the Impact and Evolution of Commit Message Quality</b> 2023 <ul style="list-style-type: none"><li>Built various machine learning classifiers (<b>Bi-LSTM</b>, <b>Bi-GRU</b>, <b>XGBoost</b>, <b>SVM</b>) to automatically evaluate commit message quality.</li><li>Considering the quality of the referenced links in the commit messages, the classifier outperformed state-of-the-art machine learning classifiers by <b>12 percentage points</b> improvement in the F1 score.</li></ul> <b>Code Readability Estimation with Large Language Models</b> 2022

- Utilized the token probability predicted by large language models (i.e., **CodeBERT**) to measure source code readability.
- Designed various features based on the token probability and conducted a **correlation analysis** between the proposed features and the human-annotated readable scores in the readability datasets.

COMPETENCES **Programming Languages** Python, Java

**Technologies and Frameworks** PyTorch, Scikit-Learn, LangChain, AutoGen, OpenAI API,  $\text{\LaTeX}$ , Github, Shell, R

**Software Engineering** Commit message quality evaluation and generation, Test smell detection and analysis, AI-generated source code detection, Code review automation, Code generation, Defect prediction, Survey, Card sorting, Open coding

**Machine Learning** Supervised Learning, Deep Learning, Large Language Models

PEER-REVIEWED PUBLICATIONS [1] Md Rakib Hossain Misu, **Jiawei Li**, Adithya Bhattiprolu, Yang Liu, Eduardo Almeida, Iftekhar Ahmed. Test Smell: A Parasitic Energy Consumer in Software Testing. *Information and Software Technology*, 2025.

[2] Hyunjae Suh, Mahan Tafreshipour, **Jiawei Li**, Adithya Bhattiprolu, Iftekhar Ahmed. An Empirical Study on Automatically Detecting AI-Generated Source Code: How Far Are We?. *Proceedings of the 47th IEEE/ACM International Conference on Software Engineering (ICSE)*, 2025.

[3] **Jiawei Li**, David Farago, Christian Petrov, Iftekhar Ahmed. Only diff is Not Enough: Generating Commit Messages Leveraging Reasoning and Action of Large Language Model. *The ACM International Conference on the Foundations of Software Engineering (FSE)*, 2024. **\*ACM SIGSOFT Distinguished Paper Award**

[4] **Jiawei Li**, Iftekhar Ahmed. Commit Message Matters: Investigating Impact and Evolution of Commit Message Quality. *Proceedings of the 45th IEEE/ACM International Conference on Software Engineering (ICSE)*, 2023.

[5] Tongjie Wang, Yaroslav Golubev, Oleg Smirnov, **Jiawei Li**, Timofey Bryksin, Iftekhar Ahmed. PyNose: A Test Smell Detector For Python. *36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2021.

[6] Jiri Gesi, **Jiawei Li**, Iftekhar Ahmed. An Empirical Examination of the Impact of Bias on Just-in-time Defect Prediction. *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*., 2021.

PAPERS IN PREPARATION

[7] EG Santana Jr, Gabriel Benjamin, Melissa Araujo, Harrison Santos, David Freitas, Eduardo Almeida, Paulo Anselmo da Neto, **Jiawei Li**, Jina Chun, Iftekhar Ahmed. Which Prompting Technique Should I Use? An Empirical Investigation of Prompting Techniques for Software Engineering Tasks. *arXiv preprint arXiv:2506.05614*, 2025.

[8] Jina Chun, Qihong Chen, **Jiawei Li**, Iftekhar Ahmed. Is Multi-Agent Debate (MAD) the Silver Bullet? An Empirical Analysis of MAD in Code Summarization and Translation. *arXiv preprint arXiv:2503.12029*, 2025.

[9] **Jiawei Li**, David Farago, Christian Petrov, Iftekhar Ahmed. Consider What Humans Consider: Optimizing Commit Message Leveraging Contexts Considered By Human. *arXiv preprint arXiv:2503.11960v2*, 2025.

[10] Qihong Chen, **Jiawei Li**, Jiecheng Deng, Jiachen Yu, Justin Tian Jin Chen, Iftekhar Ahmed. A Deep Dive Into Large Language Model Code Generation Mistakes: What and Why?. *arXiv preprint arXiv:2411.01414*, 2024.

[11] Qihong Chen, **Jiawei Li** <equal contribution>, Hyunjae Suh, Lianghao Jiang, Zheng Zhou, Jingze Chen, Jiri Gesi, Iftekhar Ahmed. Does the Order of Fine-tuning Matter and Why? *arXiv preprint arXiv:2410.02915*, 2024.

- [12] Md Rashedul Hasan, **Jiawei Li**, Iftekhar Ahmed, Hamid Bagheri. Automated Repair of Declarative Software Specifications in the Era of Large Language Models. *arXiv preprint arXiv:2310.12425*, 2023.
- [13] Jiri Gesi, Siqi Liu, **Jiawei Li**, Iftekhar Ahmed, Nachiappan Nagappan, David Lo, Eduardo Santana de Almeida, Pavneet Singh Kochhar, Lingfeng Bao. 2022. Code Smells in Machine Learning Systems. *arXiv preprint arXiv:2203.00803*, 2022.
- [14] Yaroslav Golubev, **Jiawei Li**, Viacheslav Bushev, Timofey Bryksin, Iftekhar Ahmed. Changes from the trenches: Should we automate them?. *arXiv preprint arXiv:2105.10157*, 2021.