MYEONGSOO KIM

Santa Clara, CA 95054 (+1) 678-576-5708 codingsoo93@gmail.com https://codingsoo.github.io/

SUMMARY

Applied Scientist working on coding agents, evaluation design, and LLM-guided software engineering. Designed structured action spaces that reduce token consumption by up to 38% while improving pass@1 accuracy (ACL 2026), built multi-turn evaluation benchmarks adopted org-wide to gate agent updates (NeurIPS 2025), and fine-tuned small language models achieving state-of-the-art test coverage with 80% cost reduction over baselines (FSE 2025). Earlier work applied reinforcement learning to train exploration policies for automated API testing (ASE 2023, ICSE 2025). Currently building a self-improvement loop at AWS that mines real user traffic for agent failures, reproduces them as benchmarks, and validates fixes through reproducibility and fairness gates. 16 publications at top venues (ICSE, FSE, ASE, ISSTA, NeurIPS, ACL). Interested in post-training and RL-based approaches to improving code agent capabilities.

WORK EXPERIENCE

Amazon Web Services, California, United States *Dec 2024 – Present*
Applied Scientist II (L5), AWS AI Labs — Kiro Org

- Designed CodeStruct, an AST-based structured action space for code agents that reduces token consumption by up to 38% while improving pass@1 accuracy; filed patent and published at ACL 2026.
- Building Kiro Trace, a self-improvement loop that mines live user traffic for agent failures, reproduces them as runnable benchmarks, and validates fixes through reproducibility and fairness gates before shipping.
- Built CodeAssistBench (NeurIPS 2025), a multi-turn evaluation benchmark for chat-based code assistance, adopted org-wide to gate all prompt, tool, and routing updates.
- Led agent architecture improvements for Kiro — planning workflows, tool orchestration, and prompt engineering — reducing median response latency by 14% across 1M+ production messages; two patents filed.
- Developed TrajEval, a diagnostic evaluation framework for "Coherence Collapse" — diagnosing why code agents fail after reaching the right code by measuring per-capability (search, read, write) success via oracle trajectory grounding (under review, EMNLP 2026).

IBM Research, New York, United States *May 2024 – Aug 2024*
PhD AI Research Internship with the AI for Code Team

- Automated Unit Test Generation Project: Co-developed static analysis-guided LLM test generation system combining symbolic execution with LLM-based assertion generation. Integrated retrieval-augmented generation for framework-specific patterns, leading to ICSE 2025 submission.
- Python Unit Test Generator: Designed and trained the Python component using supervised fine-tuning on 50k+ test examples, achieving 9.8% improvement in line coverage, 26.5% in branch coverage, and 22.5% in method coverage vs. Microsoft's CodaMOSA. Conducted 160-participant survey at IBM demonstrating superior effectiveness over baselines.
- Java Unit Test Generator: Evaluated Java component using mutation coverage and semantic correctness metrics, benchmarking against EvoSuite across 200+ open-source projects to assess test quality and fault-detection capability.
- Multi-Agent Integration Test Generator: Architected multi-agent system using hierarchical planning where specialized agents collaborate on API endpoint discovery (static analysis agent), dependency resolution (constraint solver agent), and test case synthesis (LLM generation agent). Designed inter-agent communication protocols for REST API testing orchestration.

Samsung, Seoul, South Korea *December 2023 – January 2024*
PhD AI Internship with the AI Team in the Fire & Marine Insurance Department

- Developed BoToPa, an LLM-based chatbot using multi-layer RAG (GPT-4 + custom BM25 tokenizer + KNN text search) to match insurance inquiries to documents; achieved 90% cost reduction and 10% accuracy improvement over the previous system.

Google, California, United States *May 2022 – August 2022*
PhD SWE Internship with the Google Chat team at Google Cloud

- Analyzed potential security issues in chat applications.
- Developed a grammar-based testing tool for internal and external chat applications which results in finding a severe security issue in Google Chat application.

IBM Research, New York, United States *June 2020 – November 2020*
PhD Research Internship with the AI for Code Team at the Thomas J. Watson Research Center

- Investigated challenges in decomposing legacy web applications using Kubernetes.
- Developed a Command Line Interface tool for validating transformation of the monolithic application to microservice application.

EDUCATION

Georgia Institute of Technology, Georgia, United States *Aug 2019 – Dec 2024*
Ph.D. in Computer Science, Advisor: Prof. Alessandro Orso

- Developed multi-agent reinforcement learning for REST API test generation, training collaborative agents with custom reward shaping to maximize code/fault coverage (ASE 2023, ISSTA 2022).
- Designed semantic graph-based state representation enabling agents to reason about resource dependencies across 15+ real-world services, achieving 34% higher branch coverage than baselines (ICSE 2025).
- Built LLM-guided test generation combining fine-tuned small language models with reinforcement learning, reducing generation cost by 80% while maintaining effectiveness (FSE 2025).
- Enhanced API documentation understanding through NLP techniques (BERT fine-tuning, semantic similarity, entity extraction), reducing invalid requests by 45% (ISSTA 2023).

Kookmin University, Seoul, South Korea *Mar 2015 – Aug 2019*
B.S. in Computer Science

PUBLICATIONS

Note: ICSE, FSE, ASE, and ISSTA are widely recognized as the top four conferences in the field of Software Engineering.

Myeongsoo Kim, Dingmin Wang, Siwei Cui, Farima Farmahinifarahani, Terry Yue Zhuo, Shweta Garg, Baishakhi Ray, Rajdeep Mukherjee, Varun Kumar. *Coherence Collapse: Diagnosing Why Code Agents Fail After Reaching the Right Code*. Under review at EMNLP 2026.

Myeongsoo Kim, Joe Hsu, Dingmin Wang, Shweta Garg, Varun Kumar, Murali Krishna Ramanathan. *CodeStruct: Code Agents over Structured Action Spaces*. In Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (ACL). 2026.

Myeongsoo Kim, Shweta Garg, Baishakhi Ray, Varun Kumar, Anoop Deoras. *CodeAssistBench (CAB): Dataset & Benchmarking for Multi-turn Chat-Based Code Assistance*. In Proceedings of NeurIPS 2025, Datasets & Benchmarks Track. 2025.

Tyler Stennett, **Myeongsoo Kim**, Saurabh Sinha, Alessandro Orso. *AutoRestTest: A Tool for Automated REST API Testing Using LLMs and MARL*. In Proceedings of the 47th IEEE/ACM International Conference on Software Engineering (ICSE), Demonstrations Track. 2025.

Myeongsoo Kim, Saurabh Sinha, Alessandro Orso. *LlamaRestTest: Effective REST API Testing with Small Language Models*. Proceedings of the ACM on Software Engineering 2.FSE (2025): 465-488.

Rangeet Pan, **Myeongsoo Kim**, Rahul Krishna, Raju Pavuluri, Saurabh Sinha. *Aster: Natural and Multi-Language Unit Test Generation with LLMs*. In Proceedings of the 47th IEEE/ACM International Conference on Software Engineering (ICSE), Software Engineering in Practice (SEIP) Track. 2025. **Distinguished Paper Award**.

Myeongsoo Kim, Saurabh Sinha, Alessandro Orso. *A Multi-Agent Approach for REST API Testing with Semantic Graphs and LLM-Driven Inputs*. In Proceedings of the 47th IEEE/ACM International Conference on Software Engineering (ICSE). 2025.

Myeongsoo Kim, Santosh Pande, Alessandro Orso. *Improving Program Debloating with 1-DU Chain Minimality*. 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion). p. 384–385.

Myeongsoo Kim, Tyler Stennett, Dhruv Shah, Saurabh Sinha, Alessandro Orso. *Leveraging Large Language Models to Improve REST API Testing*. In Proceedings of the ICSE-NIER 2024, IEEE/ACM International Conference on Software Engineering - New Ideas and Emerging Results Track (ICSE-NIER). 2024. p. 37-41.

Myeongsoo Kim, Saurabh Sinha, Alessandro Orso. *Adaptive REST API Testing with Reinforcement Learning.* In Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). 2023. p. 446-458.

Myeongsoo Kim, Davide Corradini, Michele Pasqua, Mariano Ceccato, Alessandro Orso, Saurabh Sinha, Rachel Tzoref-Brill, Mariano Ceccato. *Enhancing REST API Testing with NLP Techniques.* In Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA). 2023. p. 1232-1243.

Myeongsoo Kim, Qi Xin, Saurabh Sinha, Alessandro Orso. *Automated Test Generation for REST APIs: No Time to Rest Yet.* In Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA). 2022. p. 289-301.

Qi Xin, **Myeongsoo Kim**, Qirun Zhang, Alessandro Orso. *Subdomain-Based Generality-Aware Debloating.* In Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering (ASE). 2020. p. 224-236.

Qi Xin, **Myeongsoo Kim**, Qirun Zhang, Alessandro Orso. *Program Debloating via Stochastic Optimization.* In Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER). 2020. p. 65-68. **CS 7001 Award, GT Computing**.

Myeongsoo Kim, Changheon Song, Hyeji Kim, Deahyun Park, Yeeji Kwon, Eun Namkung, Ian G. Harris, Marcel Carlsson. *Scam Detection Assistant: Automated Protection from Scammers.* In 2019 First International Conference on Societal Automation (SA). IEEE, 2019. p. 1-8.

Elahe Paikari, JaeEun Choi, SeonKyu Kim, Sooyoung Baek, **Myeongsoo Kim**, SeungEon Lee, ChaeYeon Han, YoungJae Kim, KaHye Ahn, Chan Cheong, André van der hoek. *A chatbot for conflict detection and resolution.* In 2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering (BotSE). IEEE, 2019. pp. 29-33.

Myeongsoo Kim, Eun-Jin Im. *Ethereum Smart Contract Vulnerability Detector.* Journal of the Korea Information Science Association, 2018, pp. 1940-1942.