Gabriel Ryan

gabe@cs.columbia.edu | www.cs.columbia.edu/~gabe

Summary:

I currently work as a Senior Research Scientist at Micrsoft CoreAI, where I conduct research on improving the performance of large language models (LLMs) for software engineering tasks, including code completion, SWE agents, code review, and code retrieval. Prior to joining Microsoft, I completed my Ph.D. in Computer Science at Columbia University, where my research focused on machine learning for program analysis and software engineering.

Education:

Ph.D. Columbia University December 2023

- Ph.D. Candidate in Computer Science advised by Professor Suman Jana.

M.S. Columbia University February 2018

- M.S. in Computer Science
- Worked as Research Assistant in Intrusion Detection Lab modeling and simulating user behavior.

B.S. Swarthmore College June 2013

- B.S. Engineering, B.A. Computer Science

Experience:

Senior Research Scientist, Microsoft CoreAI Mar 2024 - Present

- Conducting research on improving the performance of large language models (LLMs) for software engineering tasks, including code completion, SWE agents, code review, and code retrieval.
- Conducted experiments on optimal post-training tasks and training data curation for GPT-4o-Copilot, the model currently used in Github Copilot.

Doctoral Researcher, Columbia Security Lab Sep 2018 - Dec 2023

- Developed novel linux kernel race prediction analysis with probabilistic and spectral learning. Published in Oakland S&P 2023.
- Developed *Proximal Gradient Analysis* (PGA), a new method for dataflow analysis that uses nonsmooth optimization methods. Implemented in LLVM. Published in USENIX Security 2021.
- Designed the *Continuous Logic Network* (CLN), a neural architecture for logical learning and reasoning, and applied it to learning program invariants for verification. Implemented library for constructing CLNs in Pytorch. Published in ICLR 2020 and PLDI 2020.

Research Intern, AWS A.I. Labs Jun 2023 - Aug 2023

- Research internship with AWS CodeWhisperer research group advised by visiting Prof. Baishahki Ray.
- Developed novel approach to regression testing with LLMs using static analysis to prompt the model to reason symbolically about program execution paths.Achieved improvements of $2\times$ coverage and $3\times$ correct test generations over baseline LLM generation approaches when evaluated on CodeGen2 and OpenAI GPT-3.5 and GPT-4 models.
- Published *Code-Aware Prompting: A study of Coverage Guided Test Generation in Regression Setting using LLM* in FSE '24.

Research Intern, Microsoft Research Jun 2021 - Aug 2021

- Research internship with MSR RiSE group advised by Sr. Pr. Scientists Todd Mytkowitz and Shuvendu Lahiri.
- Developed *TOGA: A Neural Method for Test Oracle Generation* using Transformers (CodeBERT) and a specialized grammar to generate unit tests and demonstrated a 170% improvement over prior work in bug detection.
- Published work in ICSE 2022, awarded ACM SigSoft Distinguished Paper Award.

Software Engineer, Allure Security Technology Sep 2015 - August 2018

- Developed web application for managing large scale deployments of User Behavior sensors and data loss detection using Java Spring Framework with AngularJS front end and Hibernate with Postgres and Mongodb databases. Designed and implement REST API using Swagger specification and Oauth2 authentication.
- Developed midstream network based document interception and tracking system using Squid Proxy and ICAP.

Robotics Software Consultant, 3DDataLtd May 2015 - Aug 2015

- Developed a framework for 3d Simultaneous Localization and Mapping using the Lidar Odometry and Mapping algorthim. Modified algorithm to integrate IMU Sensor data using an Extended Kalman Filter.

Radar Software Engineer, Raytheon Sep 2013 - April 2015

- Analyzed data from radar testing and provide software support for automated calibration, satellite tracking, antenna diagnostics, and maintenance prioritization software in Ada and C++.
- Earned team achievement award for completing new diagnostic capabilities ahead of schedule.

Publications:

[OOPSA 2024] *Accurate Data Race Prediction in the Linux Kernel through Sparse Fourier Learning.* G. Ryan, B. Cetin, Y. Lim, S. Jana. https://dl.acm.org/doi/pdf/10.1145/3649840

[FSE 2024] *Code-Aware Prompting: A study of Coverage Guided Test Generation in Regression Setting using LLM.* G. Ryan, S. Jain, M. Shang, S. Wang, X. Ma, M. Ramanathan, B. Ray. https://arxiv.org/pdf/2402.00097

[Oakland S&P 2023] *Precise Detection of Kernel Data Races with Probabilistic Lockset Analysis.* G. Ryan, A. Shah, D. She, S. Jana. https://cs.columbia.edu/~gabe/files/oakland2023_pla.pdf

[ICSE 2022] *TOGA: A Neural Method for Test Oracle Generation.* E. Dinella*, G. Ryan*, T. Mytkowitz, S. Lahiri. https://arxiv.org/pdf/2109.09262.pdf **(Distinguished Paper Award)**

[OSDI 2021] *DistAI: Data-Driven Automated Invariant Learning for Distributed Protocols.* J. Yao, R. Tao, R. Gu, J. Nieh, S. Jana, G. Ryan. https://www.usenix.org/system/files/osdi21-yao.pdf **(Best Paper Award)**

[USENIX Security 2021] *Fine Grained Dataflow Tracking with Proximal Gradients.* G. Ryan, A. Shah, D. She, K. Bhat, and S. Jana. https://arxiv.org/abs/1909.03461

[PLDI 2020] *Learning Nonlinear Loop Invariants with Gated Continuous Logic Networks.* J. Yao*, G. Ryan*, J. Wong*, S. Jana, and R. Gu. https://arxiv.org/abs/2003.07959

[ICLR 2020] *CLN2INV: Learning Loop Invariants with Continuous Logic Networks.* G. Ryan*, J. Wong*, Y. Jianan*, R. Gu, and S. Jana. https://arxiv.org/abs/1909.11542

[InfoVis 2018] *At a Glance: Pixel Approximate Entropy as a Measure of Line Chart Complexity.* G. Ryan, A. Mosca, R. Chang, and E. Wu. https://arxiv.org/abs/1811.03180

[Oakland S&P Workshops 2018] *Simulated User Bots: Real Time Testing of Insider Threat Detection Systems.* P. Dutta, G. Ryan, A. Zieba, and S. Stolfo. https://ieeexplore.ieee.org/document/8424654

[Oceans 2012] Oversampling MAVS for reduction of vortex-shedding velocity-sensing noise. A. Williams, G. Ryan, and T. Thwaites. https://ieeexplore.ieee.org/document/6404777

Awards:

National Defense Science and Engineering Graduate Fellowship (NDSEG). Awarded NDSEG Fellowship for proposal "Proximal Gradient Analysis for Vulnerability Detection and Defense." 2019.

NSF Graduate Research Fellowship Honorable Mention. Accorded honorable mention for proposal "Modeling and Simulating Adversarial User Behavior with Sequential VAEs." 2018.

Academic Service:

TOSEM 2024. Reviewer.

AAAI 2024. Program Committee.

TOSEM 2023. Reviewer.

AAAI 2023. Reviewer.

Teaching:

Teaching Assistant. Intrusion Detection Systems. Fall 2022.

Lecturer. Continuous Logic Networks. Spring 2021.

Teaching Assistant. Intrusion Detection Systems. Fall 2020.