Gabriel Ryan

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

Education:

Ph.D. Columbia University — Sep 2018 - Present
- Ph.D. Candidate in Computer Science with research focus on deep learning for program analysis
- Co-advised by Professors Suman Jana and Salvatore Stolfo
- Awarded NDSEG Fellowship for proposal "Proximal Gradient Analysis for Vulnerability Detection and Defense."
- Received NSF Graduate Research Fellowship Honorable Mention for proposal "Modeling and Simulating Adversarial User Behavior with Sequential VAEs."

M.S. Columbia University — Sep 2016 - December 2017
- M.S. in Computer Science (GPA 3.97)
- Conducted research in modeling and simulating user behavior for Intrusion Detection Lab.
- Selected Coursework: Security, Deep Learning, Adv. Deep Learning, Data Analytics, Data Exploration Systems, Machine Learning, Natural Language Processing, Operating Systems, Compilers, High Perf. Computing

B.S. Swarthmore College — June 2013
- B.S. Engineering, B.A. Computer Science (In Major GPA 3.83)
- Selected Coursework: Adv. Algorithms, Theory of Computation, Artificial Intelligence, Probabilistic Robotics, Computer Systems, Computer Languages, Applied Statistics, Optimization, Linear Systems Analysis

Research Experience:

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 neural transformers and a specialized grammar to automatically generate unit tests that are highly effective at finding bugs (170% improvement over any other evaluated system). Published in ICSE 2022.

Graduate Research Assistant, Columbia Security Lab — Sep 2018 - Present
- Exploring applications of Transformers and Graphical Neural Networks to guide test generation and API fuzzing advised by Prof. Suman Jana.
- Designed the *Continuous Logic Network* (CLN), a new neural architecture that directly models logic with SMT, and led a team of 3 students to implement a new loop invariant inference system (CLN2INV), that is the first to solve all 124 theoretically solvable loop invariant inference problems in the Code2Inv dataset. Published in ICLR 2020, with follow up work in PLDI 2020.
- Developed *Proximal Gradient Analysis* (PGA), a new method for dynamic software analysis that uses nonsmooth optimization methods to approximate gradients over a program. Implemented PGA as an LLVM pass and demonstrated it improved f1 accuracy by 33% over current methods in measuring dataflows to branches. Published in Usenix Security 2021.

MS Graduate Research Assistant, Columbia IDS Lab — Fall 2016 - Fall 2017
- Developed novel method for modeling and generating user behavior data with neural networks. Designed architecture combining recurrent neural networks with multi task prediction to jointly model events and times and a real-time sequence. Implemented in Pytorch and Tensorflow.
- Conducted evaluation on Simulated User Bots (SUBs), and simulated attacks against 3 user behavior models: Gaussian Mixtures, One Class SVMs, and Isolation Forests. Published in IEEE S&P Workshops 2018.

Seminar Research, Data Exploration Systems — Spring 2017 - Spring 2018
- Working in Prof. Eugene Wu's lab, developed *Pixel Approximate Entropy* (PAE), a new method for measuring visual complexity in charts using approximate entropy. Developed and conducted automated Amazon Mechanical Turk user studies demonstrating that PAE effectively predicted user's ability to perform accurately perform visual analysis tasks. Published in IEEE Infoviz 2018.

Research Assistant, Woods Hole Oceanographic Institute   January 2012
- Automated filtering and parsing of data from tow tank tests of a sub-Arctic current meter in MATLAB, setting research weeks ahead of schedule.
- Conducted frequency analysis of velocity data and developed burst sampling system that improved the signal to noise ratio by a factor of 3 while minimizing computation and radio usage.
- Published work in proceedings of Oceans 2012 conference.

Professional Experience:

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 desktop client application for automated large scale document tracker injection using Electron and AngularJS.
- Developed midstream network based document interception and tracking system using Squid Proxy and ICAP server to automatically inject trackers into suspicious documents.

Robotics Software Consultant, 3DDataLtd   May 2015 - Aug 2015
- Developed a framework for 3d Simultaneous Localization and Mapping using the Lidar Odometry and Mapping algorthm. Modified algorithm to integrate IMU Sensor data using an Extended Kalman Filter. Implemented in C++ with PCL and ROS.

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++.
- Developed additions to radar software in Ada and C++ for calibration and diagnostics.
- Earned team achievement award for completing new diagnostic capabilities ahead of schedule.

Publications:

[**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`

[**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`

[**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`

Technical Summary:

Languages: Proficient in C/C++, Java, Javascript, Python, and SQL.
Technologies: Linux, Cloud Compute, Pytorch, Tensorflow, LLVM, Qemu, Z3