Rebecca Saul (she/her/hers)

rsaul@berkeley.edu • 858-210-2504 • https://rbsaul.github.io

Education

University of California, Berkeley — Berkeley, CA

- PhD student in Computer Science advised by Professor David Wagner — Sept 2024 – present
- Focus: machine learning for computer security applications
- 2024 National Defense Science and Engineering Graduate (NDSEG) Fellow
- **Coursework:** Natural Language Processing, LLM Systems

Harvard University — Cambridge, MA

- Bachelor of Arts with Highest Honors in Mathematics, *magna cum laude*, GPA: 3.945 — May 2022
- Minor in Computer Science, Language Citation in Modern Standard Arabic
- Robert Fletcher Rogers Award for Best Undergraduate Math Talk (2022)
- **CS Coursework**: Machine Learning, Data Structures and Algorithms, Introduction to Theoretical CS, Spectral Graph Theory, Probability

Publications & Selected Writings

- **Saul, R.** Liu, C., Fleischmann, N., Zak, R.J, Micinski, K., Raff, E., Holt, J. (2024). Is Function Similarity Over-Engineered? Building a Benchmark. *NeurIPS 2024 Datasets and Benchmarks Track.*
- Liu, C.[*], **Saul, R.**[*], Sun, Y., Raff, E., Fuchs, M., Southard Pantano, T., Holt, J., Micinski, K. (2024). Assemblage: Automatic Binary Dataset Construction for Machine Learning. *NeurIPS 2024 Datasets and Benchmarks Track.*
- **Saul, R.,** Alam, M.M., Hurwitz, J., Raff, E., Oates, T., Holt, J. (2023). Lempel-Ziv Networks. *Proceedings on "I Can't Believe It's Not Better! - Understanding Deep Learning Through Empirical Falsification" at NeurIPS 2022 Workshops*, in *Proceedings of Machine Learning Research* 187:1-11.
- Saul, R. (2022). Efficient Factoring and the Number Field Sieve [Senior thesis]. Harvard College.

Work Experience

Booz Allen Hamilton — Annapolis Junction, MD

Lead Scientist; Machine Learning Researcher — Apr. 2024 – Aug. 2024

Senior Consultant — Aug. 2022 – Apr. 2024

Responsible for exploring existing ML techniques, creating new algorithms, mentoring junior colleagues, and contributing to the research community through peer-reviewed papers, conference presentations, and open-source projects.

- Trained a state-of-the-art semantic binary function similarity convolutional neural network model (known as Reverse Engineering Function Search – REFuSE) for static malware analysis. Trained the model directly on raw function bytes via triplet learning with the JAX ML library.
- Developed and maintained the REFuSe codebase, which contains over 5000 lines of code, including unit tests and extensive documentation.
- Presented REFuSe at the 2024 Malware Technical Exchange Meeting (MTEM) to over 250 individuals.
- Guided the technical development, content and format of ASSEMBLAGE, a machine learning dataset of over 1M benign executables designed for binary analysis tasks such as reverse engineering, malware analysis, and vulnerability detection.
- Created a multi-modal dataset of 1.1M functions, represented by raw bytes, disassembly, decompilation, source code, and natural language description, to train large language models for binary analysis.
- Partnered with researchers at Sandia and Lawrence Livermore National Labs (LLNL) on common ML and cybersecurity efforts.

[*] denotes equal contribution

- Integrated an ML analytic for malware detection into a Java-based file analysis system in a short deployment window.
- Received the *Booz Allen Hamilton Passionate Service Award* for exceeding expectations on a "complex technical task" for a "demanding client".
- Provided technical support on projects including Neural Ordinary Differential Equations, Hopfield Networks, Large Language Model for Understanding Binaries, and Best Subset Selection for Logistic Regression.
- Coached junior contractors and government interns to create a database of crypt libraries used by malware authors.

Booz Allen Hamilton Annapolis Junction, MD
Machine Learning Research Intern (Remote) June 2021 – Aug. 2021

- Researched developments in continuous and differentiable associative memories
- Implemented a new Recurrent Neural Net architecture utilizing said associative memories in Python
- Drafted a paper detailing the project's conclusions which I presented at NeurIPS' 2022 I Can't Believe It's Not Better Empirical Falsification Workshop. The paper was subsequently published in a 2023 edition of *Proceedings of Machine Learning Research*

Harvard College Writing Center Cambridge, MA
Peer Tutor Sept. 2019 – May 2022

- Worked one-on-one with Harvard undergraduates on writing assignments in any discipline, teaching writing strategies with emphasis in the areas of academic argument, essay structure, and overall clarity
- Assisted with orientation and training of new tutors each semester
- Nominated to the position by expository writing professor

Skills & Interests

Technical: Python [incl. PyTorch and JAX], Java, LaTeX, SQL, Ghidra, Diffrax, git (proficient);
C, SAGE, Mathematica, SAT solvers, Docker, IDA Pro (familiar)
Volunteer: A-OK Mentoring and Tutoring for elementary and middle schoolers in Howard County
Interests: Soccer, basketball, piano, board games, legos, constitutional law, politics and public policy