Marina Zhang

Email: marinazhang2000@gmail.com
Website: marinazhang.github.io

Education

2017 - 2021 **Massachusetts Institute of Technology (MIT)**, *Cambridge, MA*
B.S. in Computer Science and Engineering, B.S. in Mathematics, Minor in Economics
GPA: 5.0/5.0
Relevant Coursework: Machine Learning, Computer Vision, Computational Cognitive Science, Cryptography, Algorithms, Information Theory, Statistics, Real Analysis

2014 - 2017 **Garnet Valley High School**, *Garnet Valley, PA*
Class Rank: 1 (out of 410)

Experience

Jun 2021 - Present **Google**, Security & Anti-Abuse Research Team
Software Engineer II, advised by Elie Bursztein

- Robust and Scalable NLP Models and Text Embeddings
  - Researching and building adversarially robust, multilingual, and efficient text models and embeddings using novel metric learning techniques, including RETVec [2] and RETSim [1]. Core contributor to the open-source Python packages (RETVec, UniSim, and TensorFlow Similarity) corresponding to the research.
  - Deploying my research in practice for security and anti-abuse use cases. Launched protections in numerous Google products (Gmail, YouTube, Workspace) for clustering and classifying abusive content such as malware, spam and phishing campaigns, e.g. blocking billions of spam emails in Gmail daily.
- Applied Machine Learning for Security Applications
  - Investigating novel applications of deep learning for security use cases, including large-scale malware and phishing detection, malicious URL classification for end-to-end encrypted environments, obfuscated code detection, and side-channel attacks on cryptographic hardware [3].
- Security and Reliability of LLMs
  - Conducting research on the security vulnerabilities and robustness of LLMs, including on-device text safety protections (Google I/O 2023 demo and launched for Android devices), defenses against emerging threats, and data-centric approaches to reduce privacy risk and improve generalization capabilities.

Sep 2020 - Jun 2021 **University of Pennsylvania**, Department of Neuroscience
Research Assistant, advised by Prof. Wenqin Luo

- Designed the convolutional recurrent neural network (CRNN) used for Scratch-AID, a deep learning-based tool which can automatically identify and quantify mouse scratching behavioral patterns from raw video footage. Paper published in journal (eLife) [5].
- Investigated the neurobiological relationship between behavioral states and breathing patterns using clustering techniques; built a classifier capable of distinguishing between 9 different behavioral states in rodents using breathing recordings. Paper published in journal (iScience) [4].

Jun 2020 - Aug 2020 **Google**, Security & Anti-Abuse Research Team
Software Engineer Intern, advised by Elie Bursztein

- Built a deep-learning model training framework for security research, which supported state-of-the-art model training techniques including hyperparameter tuning, semi-supervised labeling, and transfer learning.
- Designed and wrote a TensorFlow/Keras package which automates searching for and applying the best data augmentation policies during model training.

Sep 2019 - Dec 2019 **MIT CSAIL**, Medical Vision Group
Research Assistant, advised by Prof. Polina Golland

- Built an ML model for automatically quantifying the severity of pulmonary edema from patients' x-ray images and radiology reports, helping to improve clinicians' abilities to provide more accurate and personalized treatment plans for heart failure patients.

Jun 2019 - Aug 2019 **Microsoft**, Edge Browser Experiences Team
Software Engineer Intern

- Designed and added the quick-access Favorites toolbar button and drop-down menu in Microsoft Edge. This feature was shipped to all users.

Sep 2017 - Jun 2018 **MIT**, Department of Economics
Research Assistant, advised by Prof. Daron Acemoglu

- Investigated historical trends of technological change and innovation using statistical methods applied to U.S. agriculture data and economic records.

Papers

[1] **Marina Zhang**, Owen Vallis, Aysegul Bumin, Tanay Vakharia, and Elie Bursztein. "RETSim: Resilient and Efficient Text Similarity." In *International Conference on Learning Representations (ICLR) 2024.* (To appear) [Link]

[2] Elie Bursztein, **Marina Zhang**, Owen Vallis, Xinyu Jia, and Alexey Kurakin. "RETVec: Resilient and Efficient Text Vectorizer." In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. [Link]

[3] Elie Bursztein, Luca Invernizzi, Karel Král, Daniel Moghimi, Jean-Michel Picod, and **Marina Zhang**. "Generic Attacks against Cryptographic Hardware through Long-Range Deep Learning." Under review at the *Conference on Cryptographic Hardware and Embedded Systems (CHES) 2024.* [Link] *(alphabetical authorship)*

[4] Emma Janke, **Marina Zhang**, Sang Eun Ryu, Janardhan Bhattarai, Mary Schreck, Andrew Moberly, Wenqin Luo, Long Ding, Daniel Wesson, and Minghong Ma. "Machine Learning-based Clustering and Classification of Mouse Behaviors via Respiratory Patterns." In *iScience (Cell Press) Vol. 25*, 2022. [Link]

[5] Huasheng Yu, Jingwei Xiong, Adam Yongxin Ye, Suna Li Cranfill, Tariq Cannonier, Mayank Gautam, **Marina Zhang**, Rayan Bilal, Jong-Eun Park, Yuji Xue, Vidhur Polam, Zora Vujovic, Daniel Dai, William Ong, Jasper Ip, Amanda Hsieh, Nour Mimouni, Alejandra Lozada, Medhini Sosale, Alex Ahn, Minghong Ma, Long Ding, Javier Arsuaga, and Wenqin Luo. "Scratch-AID: A Deep-learning Based System for Automatic Detection of Mouse Scratching Behavior with High Accuracy." In *eLife Vol. 11*, 2022. [Link]

Open-Source Projects

2023 - Present **UniSim: Universal Similarity**, *Main Contributor*
UniSim is a Python package for efficient similarity computation, fuzzy matching, and data clustering using similarity embeddings [1].

2022 - Present **RETVec: Resilient and Efficient Text Vectorizer**, *Main Contributor*
The RETVec package offers an easy-to-use TensorFlow/Keras API for RETVec, an efficient, multilingual, and robust text vectorizer from our research [2].

2021 - Present **TensorFlow Similarity: Metric Learning for Humans**, *Contributor*
TensorFlow Similarity is a TensorFlow library for similarity learning, including algorithms and models for self-supervised learning, metric learning, and contrastive learning.

Awards & Honors

2020 Tau Beta Pi (TBP) Honor Society

2020 IEEE Eta Kappa Nu (IEEE-HKN) Honor Society

2020	MIT EECS Undergraduate Research and Innovation Scholar
2018, 2019	2x ITA Scholar-Athlete Award
2018, 2019	2x NCAA Championship Elite Eight (MIT Women's Tennis)
2017	National Merit Scholarship Winner
2017	National AP Scholar

Activities

2022 - Present **Google Intern Host**
Hosted/co-hosted two Research Scientist interns and two SWE interns at Google

2020 - 2021 **MIT IEEE-HKN Tutor for EECS**
Tutor for 6.009 Fundamentals of Programming and 6.006 Introduction to Algorithms

2019 - 2020 **MIT xFair Committee**
Organizational committee for xFair, MIT's largest student-run career fair and tech expo

2019 - 2020 **MIT UA Innovation Committee**
Worked on projects to improve student life, study spaces, and mental health at MIT

2017 - 2019 **MIT Varsity Tennis Team**
2x NCAA Elite 8; 2x ITA Scholar-Athlete; 2x NEWMAC First Team All-Conference