Dennis Jacob

djacob18@berkeley.edu

Website: https://djapp18.github.io/
GitHub: https://github.com/djapp18

EDUCATION

2025 - Present **University of California, Berkeley**, Berkeley, California
PhD Student in Electrical Engineering and Computer Sciences | Advisor: Prof. David Wagner

2020 - 2024 **Princeton University**, Princeton, New Jersey
B.S.E. in Electrical and Computer Engineering (*magna cum laude*)
Cumulative GPA: 3.92 | Certificate in Applied and Computational Mathematics

RESEARCH/WORK EXPERIENCE

Aug. 2024 – Aug. 2025 **University of California, Berkeley**, Berkeley, California
Research Assistant under Prof. David Wagner
Research on trustworthy machine learning, trustworthy AI, and robustness for large language models (LLMs). Identified ideas for research directions, developed novel techniques for robustness of LLMs, implemented candidate techniques, communicated results via research papers, etc.

Summer 2023 **Princeton University**, Princeton, New Jersey
Research Intern under Prof. Prateek Mittal
Research on adversarial machine learning (ML). Proposed a method for designing a certifiably robust defense for multi-label classifiers against the adversarial patch threat model. Demonstrated non-trivial robustness and clean performance on the MS-COCO dataset.

Summer 2022 **Princeton University**, Princeton, New Jersey
Research Intern under Prof. Sharad Malik
Research on hardware verification methods. Modeled components of the NVDLA machine learning accelerator for convolutional neural networks. Used ILAng methodology to create abstractions of hardware design.

Summer 2021 **Corning Incorporated**, Corning, New York
Research Intern
Designed, developed, and implemented a Raspberry PI-based control system for cellular ceramic filter testing in diesel engine pollution control applications. Additionally improved legacy MATLAB code through GUI development, and designed a HMI + PLC programming interface for a burner rig testing suite. Documented the work via Corning Internal Research Reports.

Summer 2019 **Corning Incorporated**, Corning, New York
Highschool Research Intern
Developed and optimized a convolutional neural network (CNN) based tool for cellular ceramic manufacturing process improvement. Resulted in a Corning Internal Research Report.

RESEARCH COLLABORATIONS

June 2023 – Present **Karlsruhe Institute of Technology**, Karlsruhe, Germany
Research Collaborator with Dr. Sven Banisch
Investigating the causes and structure of polarization in online platforms. We leverage agent-based modeling (ABM) to model individual preferences and a combination of reinforcement learning (RL) and dynamical systems techniques to understand underlying opinion dynamics.

RESEARCH PUBLICATIONS

2025 **JailbreaksOverTime: Detecting Jailbreak Attacks Under Distribution Shift**
Julien Piet, Xiao Huang, Dennis Jacob*, Annabella Chow, Maha Alrashed, Geng Zhao, Zhanhao Hu, Chawin Sitawarin, Basel Alomair, David Wagner*, 18th ACM Workshop on Artificial Intelligence and Security (AISec '25, co-located with CCS '25)

2025 **PromptShield: Deployable Detection for Prompt Injection Attacks**
Dennis Jacob*, *Hend Alzahrani*, Zhanhao Hu, Basel Alomair, and David Wagner*, The 15th ACM Conference on Data and Application Security and Privacy (ACM CODASPY 2025) (* denotes equal contribution)

2025 **PatchDEMUX: A Certifiably Robust Framework for Multi-label Classifiers Against Adversarial Patches**
Dennis Jacob*, Chong Xiang, and Prateek Mittal*, 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025)

2024 **A dynamical model of platform choice and online segregation**
Sven Banisch, Dennis Jacob*, Tom Willaert, and Eckehard Olbrich*, Preprint (arXiv).

2024 **WIP: Towards a Certifiably Robust Defense for Multi-label Classifiers Against Adversarial Patches**
Dennis Jacob, *Chong Xiang, and Prateek Mittal*, Workshop on Artificial Intelligence System with Confidential Computing (AISCC 2024, co-located with NDSS Symposium 2024), *Distinguished Paper Award*

2023 **Polarization in Social Media: A Virtual Worlds-Based Approach**
Dennis Jacob *and Sven Banisch*, Journal of Artificial Societies and Social Simulation (JASSS) 26 (3) 11.

PATENTS

2024 **US11969051B2: Internet connected adjustable structural support and cushioning system for footwear** *(method patent)*
Dennis George Jacob (April 30, 2024).

2022 **US11464286B2: Internet connected adjustable structural support and cushioning system for footwear** *(system patent)*
Dennis George Jacob (Oct. 11, 2022).

HONORS and AWARDS

2024 G. David Forney, Jr. Prize (Princeton University): Outstanding Senior Thesis in ECE
2024 Sigma Xi Honor Society
2024 Tau Beta Pi Honor Society
2022 Shapiro Prize for Academic Excellence (Princeton University)
2020 - 2024 National Merit Scholarship award
2019 National Finalist in Young Entrepreneurs Academy (YEA!) competition

ACADEMIC SERVICES

Peer Reviewer, Journal of Artificial Societies and Social Simulation, Journal of Computational Social Science

TEACHING and MENTORING

Spring 2023 **Teaching Assistant** for *ECE 432: Information Security* (Princeton University): held weekly office hours and graded homework assignments

Fall 2022 **Teaching Assistant** for *ECE 206: Contemporary Logic Design* (Princeton University): held weekly office hours

Fall 2021 **Course Development Assistant** for *COS 324: Introduction to Machine Learning* (Princeton University): co-wrote lecture notes available at https://princeton-introml.github.io/index.html

LEADERSHIP

2023 - 2024 **Colonial Club** (Princeton University): Appointed officer of Colonial Club, one of the Princeton eating clubs. Helped plan social events, recruit members, and arrange weekly orders of food and beverages.

2021 - 2024 **Hoagie Club** (Princeton University): Vice President and founding member of Hoagie Club, a student-run software developer group. Co-led development of HoagieStuff, the exchange platform for Princeton students.

2019 **bAIR Technologies**: Founder of bAIR Technologies, an IoT technology start up in association with the YEA! Program. Invented and developed an internet-connected smart sole that can be adjusted for custom comfort and support; technology covered by 2 US patents (US11464286B2 and US11969051B2).