

# Chawin Sitawarin

Soda Hall, UC Berkeley  
Berkeley, CA 94709  
☎ 510-423-2880  
✉ [chawins@berkeley.edu](mailto:chawins@berkeley.edu)  
🌐 <https://chawins.github.io/>

## Education

- 2018–2023 **PhD in Computer Science**, *UC Berkeley*, Berkeley CA.  
(tentative) Advisor: Professor David Wagner | GPA 3.86
- 2014–2018 **BSE in Electrical Engineering (High Honor)**, *Princeton University*, Princeton NJ.  
Cumulative GPA: 3.90, Departmental GPA: 3.95 | Certificate in Applications of Computing

## Research Interests

I am broadly interested in the intersection between ML and computer security where most of my works focus on adversarial robustness. Recently, I have been excited about the new security implications of foundation models with a natural language interface. My research goal is to make ML models safe and secure in practice without compromising their utility.

## Publications

- 2023 **SPDER: Semiperiodic Damping-Enabled Object Representation**, *K. Shah, C. Sitawarin*, Preprint (under submission), [paper](#).
- 2023 **REAP: A Large-Scale Realistic Adversarial Patch Benchmark**, *N. Hingun\*, C. Sitawarin\*, J. Li, D. Wagner*, Preprint (under submission), [paper](#), [code](#).
- 2023 **Preprocessors Matter! Realistic Decision-Based Attacks on Machine Learning Systems**, *C. Sitawarin, F. Tramèr, N. Carlini*, ICML 2023 (poster), [paper](#), [code](#).
- 2023 **Part-Based Models Improve Adversarial Robustness**, *C. Sitawarin, K. Pongmala, Y. Chen, N. Carlini, D. Wagner*, ICLR 2023 (poster), [paper](#), [code](#).
- 2023 **Short: Certifiably Robust Perception against Adversarial Patch Attacks: A Survey**, *C. Xiang, C. Sitawarin, T. Wu, P. Mittal*, 1st Symposium on Vehicle Security and Privacy (NDSS 2023), Best Short/WIP Paper Award Runner-Up, [paper](#), [code](#).
- 2022 **Demystifying the Adversarial Robustness of Random Transformation Defenses**, *C. Sitawarin, Z. Golan-Strieb, D. Wagner*, ICML 2022 (short presentation) and AAAI-22 AdvML Workshop (Best Paper), [paper](#), [code](#).
- 2021 **Adversarial Examples for  $k$ -Nearest Neighbor Classifiers Based on Higher-Order Voronoi Diagrams**, *C. Sitawarin, E. M. Kornaropoulos, D. Song, D. Wagner*, NeurIPS 2021 (poster), [paper](#), [code](#).
- 2021 **Improving the Accuracy-Robustness Trade-Off for Dual-Domain Adversarial Training**, *C. Sitawarin, A. Sridhar, D. Wagner*, Workshop on Uncertainty & Robustness in Deep Learning (ICML 2021), [paper](#), [code](#).
- 2021 **Mitigating Adversarial Training Instability with Batch Normalization**, *A. Sridhar, C. Sitawarin, D. Wagner*, Workshop on Security and Safety in Machine Learning Systems (ICLR 2021), [paper](#).
- 2021 **SAT: Improving Adversarial Training via Curriculum-Based Loss Smoothing**, *C. Sitawarin, S. Chakraborty, D. Wagner*, AISec 2021 (co-located with CCS), [paper](#).
- 2020 **Minimum-Norm Adversarial Examples on  $k$ -NN and  $k$ -NN-Based Models**, *C. Sitawarin, D. Wagner*, Deep Learning and Security Workshop (IEEE S&P 2020), [paper](#).
- 2019 **Analyzing the Robustness of Open-World Machine Learning**, *V. Schwag, A. N. Bhagoji, L. Song, C. Sitawarin, D. Cullina, M. Chiang, and P. Mittal*, AISec 2019 (co-located with CCS), [paper](#).
- 2019 **Defending Against Adversarial Examples with K-Nearest Neighbor**, *C. Sitawarin, D. Wagner*, Preprint, [arXiv:1906.09525](https://arxiv.org/abs/1906.09525).

- 2018 **On the Robustness of Deep k-Nearest Neighbors**, C. Sitawarin, D. Wagner, Deep Learning and Security Workshop (IEEE S&P 2019), [paper](#).
- 2018 **Not All Pixels are Born Equal: An Analysis of Evasion Attacks under Locality Constraints**, V. Sehwag, C. Sitawarin, A. N. Bhagoji, A. Mosenia, M. Chiang, P. Mittal, CCS 2018 Poster, [paper](#).
- 2018 **DARTS: Deceiving Autonomous Cars with Toxic Signs**, C. Sitawarin, A. N. Bhagoji, A. Mosenia, M. Chiang, P. Mittal, Preprint, [arXiv:1802.06430](#).
- 2018 **Rogue signs: Deceiving Traffic Sign Recognition with Malicious Ads and Logos**, C. Sitawarin, A. N. Bhagoji, A. Mosenia, M. Chiang, P. Mittal, Deep Learning and Security Workshop (IEEE S&P 2018), [paper](#).
- 2018 **Enhancing Robustness of Machine Learning System via Data Transformations**, A. N. Bhagoji, D. Cullina, C. Sitawarin, P. Mittal, CISS 2018, [paper](#).
- 2017 **Beyond Grand Theft Auto V for Training, Testing and Enhancing Deep Learning in Self Driving Cars**, M. A. Martinez, C. Sitawarin, K. Finch, L. Meincke, A. Yablonski, A. Kornhauser, Preprint, [arXiv:1712.01397](#).

## Other Experiences

- Summer 2022 **Google**, *Sunnyvale CA*, Research Intern.  
Evaluate and mitigate machine learning security risks in a practical setting where a pair of public client-side and secret server-side models is deployed for a malware detection task. Hosted by Ali Zand and David Tao.
- Fall 2021 - **Google Brain**, *Remote*, Student Researcher (part-time).
- Spring 2022 Developed threat model and appropriate evaluation for adversarial robustness in new and practical settings (e.g., dynamic models, black-box model recovery). Hosted by Nicholas Carlini.
- Summer 2021 **Nokia Bell Labs**, *Remote*, Research Intern.  
Investigated relationships between causality and robustness in machine learning, focusing on leveraging causal relationships to improve robustness and generalization to unseen attacks/corruptions. Hosted by Anwar Walid.
- Fall 2020 & **EECS Department, UC Berkeley**, *Berkeley CA*, Graduate Student Instructor.  
Spring 2023 CS189/289A: Introduction to Machine Learning.
- Summer 2019 **IBM Research**, *Yorktown Heights NY*, Research Intern.  
Studied the effectiveness of existing defenses against adversarial examples from a perspective of concentration bound and improved adversarial training through optimization techniques. Hosted by Supriyo Chakraborty.

## Awards & Honors

- |           |  |  |
|-----------|--|--|
| 2022      | <b>Google-BAIR Commons Project</b>               | <i>Research grant</i>  |
| 2021-2022 | <b>Center for Long-Term Cybersecurity (CLTC)</b> | <i>Research grant</i>  |
| 2021      | <b>Microsoft-BAIR Commons Project</b>            | <i>Research grant</i>  |
| 2018      | <b>Phi Beta Kappa</b>                            | <i>Academic Honor Society</i>  |
| 2018      | <b>Sigma Xi</b>                                  | <i>Scientific Research Honor Society</i>   |
| 2017      | <b>The P. Michael Lion III Fund</b>              | <i>Summer research funding for Princeton engineering students</i>                          |
| 2016      | <b>Tau Beta Pi</b>                               | <i>Engineering Honor Society</i>   |
| 2016      | <b>Shapiro Prize for Academic Excellence</b>     | <i>Academic award at Princeton University</i>  |
| 2013      | <b>King's Scholarship</b>                        | <i>Prestigious scholarship awarded by Thai government for pursuing a bachelor's degree</i> |

## Activities and Services

- Program Committee**, *AISeC 2022, 2023*.
- Reviewer**, *ICML 2022 | NeurIPS 2022, 2023 | BANDS (ICLR workshop) 2023*.
- 2019–present **DARE: Diversifying Access to Research in Engineering**, *Mentor*, Mentored under-represented students in CS on multiple research projects..
- 2018–2020 **CSGSA**, *Treasurer*, Computer Science Graduate Student Assembly at UC Berkeley.

---

## Other Publications

- 2018 **Enhancing Robustness of Classifiers Against Adversarial Examples**, Undergraduate Thesis, Advisor: Professor Peter Ramadge.
- 2018 **Inverse-Designed Photonic Fibers and Metasurfaces for Nonlinear Frequency Conversion**, *C. Sitawarin, Z. Lin, W. Jin and A. W. Rodriguez*, Photonics Research Vol. 6, Issue 5, [paper](#).
- 2016 **Inverse-designed nonlinear nanophotonic structures: Enhanced frequency conversion at the nano scale**, *Z. Lin, C. Sitawarin, M. Lončar, A. W. Rodriguez*, Conference on Lasers and Electro-Optics (CLEO) 2016, [paper](#).