Arjun Bhagoji

Research Scientist
Department of Computer Science
University of Chicago
Chicago, IL 60637

(+1)-609-558-2325
abhagoji@uchicago.edu
arjunbhagoji.github.io

Research Interests

My research has two broad directions. The first focuses on the reliability of machine learning (ML) systems, based on analyses of the underlying algorithms. I am interested in understanding how intelligent, and possibly malicious agents, can influence machine learning algorithms to achieve their own goals. I use this understanding to guide the development of robust algorithms and find the fundamental limits of learning in the presence of adversaries. The second direction concerns reliable ML for society. I develop reliable ML systems for problems of societal interest such as nation-state censorship detection and internet access, and analyze the impact of ML systems on issues such as content moderation.

Education

Program	**Institution**	**Years**
Ph.D., Electrical and Computer Engineering	Princeton University Princeton, NJ	2015 – 2020
B.Tech. (Honors) and M.Tech., Electrical Engineering (minor in Physics)	Indian Institute of Technology Madras Chennai, India	2010 – 2015

Experience

- **Research Scientist** in the Department of Computer Science at the University of Chicago [Feb. 2023 - Present]
- **Postdoctoral Scholar** in the Department of Computer Science at the University of Chicago working with Ben Zhao and Nick Feamster on reliable machine learning [Sept. 2020 - Feb. 2023]
- **Graduate Research Assistant** at Princeton University advised by Prof. Prateek Mittal working on the security of machine learning systems [Feb. 2016 - August 2020]
- **Summer Research Intern** at the I.B.M. T.J. Watson Research Center with Dr. Supriyo Chakraborty studying the robustness of distributed learning systems [May - Sept. 2018]
- **Visiting student** at UC Berkeley with Prof. Dawn Song working on practical and theoretical limits of black-box attacks on machine learning systems [June - Sept. 2017]

Awards & Achievements

- **Principal Investigator on a $150k award from C3.ai** for "Fundamental Limits on the Robustness of Supervised Machine Learning Algorithms" with Ben Zhao and Daniel Cullina
- Selected as a **UChicago Rising Star in Data Science 2021**
- Finalist for the **Bede Liu Best Dissertation Award 2020 in the ECE Department, Princeton University**
- Recipient of the **Yan Huo *94 Graduate Fellowship in Electrical Engineering 2019**
- Recipient of the **Siemens FutureMakers Fellowship in Machine Learning 2018**

- Finalist for the **Bell Labs Prize 2017 (Top 9 of over 300 participating teams)**
- Recipient of the First Year Fellowship (2015-2016) for graduate studies at Princeton University
- Awarded the DAAD WISE Scholarship 2013 for a summer internship in Germany

Publications

Total citations: 8113, h-index: 16, i10-index: 19 (Google Scholar: https://rb.gy/5r6bya)

Working Papers

- **A. Bhagoji***, Z. Sarwar*, V. Tran*, N. Feamster and B. Y. Zhao, "Enola: Towards Effective External Data Augmentation", To be submitted to *Transactions on Machine Learning Research*
- **A. Bhagoji**, Y. Wang, D. Cullina and B. Y. Zhao, "Bounding the Robustness of Fixed Feature Extractors to Test-time Adversaries", To be submitted to *Transactions on Machine Learning Research*
- H. Li, **A. Bhagoji**, H. Zheng and B. Y. Zhao, "Can Backdoor Attacks Survive Time-Varying Models?", To be submitted to *ICML 2024*

Book Chapters

- **A. Bhagoji** and S. Chakraborty, "Assessing vulnerabilities and securing federated learning", *Federated Learning: Theory and Practice*, doi
- **A. Bhagoji** and P. Shirani, "Adversarial Attacks for Anomaly Detection", *Springer Encyclopedia of Machine Learning and Data Science*, doi

Papers

- B. Schaffner*, **A. Bhagoji***, S. Cheng, J. Mei, J. Shen, G. Wang, M. Chetty, N. Feamster, G. Lakier, C. Tan, "Community Guidelines Make this the Best Party on the Internet: An In-Depth Study of Online Platforms' Content Moderation Policies", *CHI 2024*
- W. Ding, **A. Bhagoji**, H. Zheng and B. Y. Zhao, "Towards Scalable and Robust Model Versioning", *IEEE SaTML 2024*, arXiv:2401.09574
- X. Jiang, S. Liu, A. Gember-Jacobson, **A. Bhagoji**, P. Schmitt, F. Bronzino, N. Feamster, "NetDiffusion: Network Data Augmentation Through Protocol-Constrained Traffic Generation", Accepted with Shepherding for *Proceedings of the ACM on Measurement and Analysis of Computing Systems (POMACS)*, 2024, arXiv:2310.08543
- S. Dai*, W. Ding*, **A. Bhagoji**, D. Cullina, B.Y. Zhao, H. Zheng, P. Mittal , "Characterizing the Optimal 0-1 Loss for Multi-class Classification with a Test-time Attacker", **Spotlight** at *NeurIPS 2023*, arXiv:2302.10722
- S. Liu, F. Bronzino, P. Schmitt, **A. Bhagoji**, N. Feamster, H. G. Crespo, T. Coyle, B. Ward, "LEAF: Navigating Concept Drift in Cellular Networks", *Proceedings of the ACM on Networking (PACMNET)*, 2023, arXiv:2109.03011
- J. Brown*, V. Tran*, X. Jiang*, **A. Bhagoji**, N.P. Hoang, N. Feamster, P. Mittal and V. Yegneswaran, "Exploring the Feasibility of Machine Learning-Driven DNS Censorship Detection", *KDD 2023*, arXiv:2302.02031
- C. Cianfarani*, **A. Bhagoji***, V. Sehwag*, B. Zhao, H. Zheng, P. Mittal, "Understanding robust learning through the lens of representation similarities", *NeurIPS 2022*, arXiv:2206.09868
- E. Wenger, R. Bhattacharjee, **A. Bhagoji**, J. Passananti, E. Andere, H. Zheng and B. Y. Zhao, "Finding Naturally Occurring Physical Backdoors in Image Datasets", *NeurIPS 2022*, arXiv:2206.10673
- S. Shawn, **A. Bhagoji**, H. Zheng and B. Y. Zhao, "Traceback of Data Poisoning Attacks in Neural Networks", *USENIX Security 2022*, arXiv:2110.06904
- A. Panda, S. Mahloujifar, **A. Bhagoji**, S. Chakraborty and P. Mittal, "SparseFed: Mitigating Model Poisoning Attacks in Federated Learning with Sparsification", *AISTATS 2022*, arXiv:2112.06274

- S. Shawn, **A. Bhagoji**, H. Zheng and B. Y. Zhao, “A Real-time Defense against Website Fingerprinting Attacks”, *AISec 2021*, arXiv:2102.04291
- C. Xiang, **A. Bhagoji**, V. Sehwag and P. Mittal, “PatchGuard: A Provable Robust Defense against Adversarial Patches via Small Receptive Fields and Masking”, *USENIX Security 2021*, arXiv:2005.10884
- **A. Bhagoji***, D. Cullina*, V. Sehwag and P. Mittal, “Lower Bounds on Cross-Entropy Loss in the Presence of Test-time Adversaries”, *ICML 2021*, arXiv:2104.08382
- E. Wenger, J. Passananti, **A. Bhagoji**, Y. Yao, H. Zheng and B. Y. Zhao, “Backdoor Attacks on Facial Recognition in the Physical World”, *CVPR 2021*, arXiv:2006.14580
- **A. Bhagoji***, D. Cullina* and P. Mittal, “Lower Bounds on Adversarial Robustness from Optimal Transport”, *NeurIPS 2019*, arXiv:1909.12272
- P. Kairouz, H. B. McMahan, **A. Bhagoji**, et.al., “Advances and Open Problems in Federated Learning”, *Foundations and Trends in Machine Learning (FnTML)*, 2021, arXiv:1912.04977
- V. Sehwag*, **A. Bhagoji***, L. Song*, C. Sitawarin, D. Cullina, A. Mosenia, P. Mittal and M. Chiang, “Analyzing the Robustness of Open-World Machine Learning”, *AISec 2019*, arXiv:1905.01726
- **A. Bhagoji**, S. Chakraborty, P. Mittal and S. Calo, “Analyzing Federated Learning through an Adversarial Lens”, *ICML 2019*, arXiv:1811.12470
- D. Cullina*, **A. Bhagoji*** and P. Mittal , “PAC-learning in the presence of evasion adversaries”, *NeurIPS 2018*, arXiv:1806.01471
- **A. Bhagoji**, W. He, B. Li and D. Song, “Practical Black-box Attacks on Deep Neural Networks using Efficient Query Mechanisms”, *ECCV 2018*, CVF Open Access
- **A. Bhagoji**, D. Cullina, C. Sitawarin and P. Mittal , “Enhancing robustness of machine learning systems via data transformations”, *CISS 2018*, doi
- **A. Bhagoji** and P. Sarvepalli, “Equivalence of 2D color codes (without translational symmetry) to surface codes”, *ISIT 2015*, doi

Workshop papers

- **A. Bhagoji**, D. Cullina and B. Y. Zhao, “Lower Bounds on the Robustness of Fixed Feature Extractors to Test-time Adversaries”, *2nd New Frontiers in Adversarial Machine Learning 2023 (AdvML Frontiers at ICML)*
- V. Sehwag, **A. Bhagoji**, C. Sitawarin, A. Mosenia, M. Chiang and P. Mittal, “Not all pixels are born equal: An analysis of evasion attacks under locality constraints", *ACM CCS 2018*
- C. Sitawarin*, **A. Bhagoji***, A. Mosenia, M. Chiang and P. Mittal, “Rogue Signs: Deceiving Traffic Sign Recognition with Malicious Ads and Logos”, *1st Deep Learning and Security Workshop (co-located with IEEE S&P)*, 2018, arXiv:1801.02780
- N. Sivadas, **A. Bhagoji** et. al., “A Nanosatellite Mission to Study Charged Particle Precipitation from the Van Allen Radiation Belts caused due to Seismo-Electromagnetic Emissions”, *5th Nano-Satellite Symposium*, 2013, arXiv:1411.6034

Theses

- **A. Bhagoji**, “The Role of Data Geometry in Adversarial Machine Learning”, *Ph. D. Thesis, Department of Electrical and Computer Engineering, Princeton University*, 2020, ProQuest
- **A. Bhagoji**, “Equivalence of color codes and surface codes”, *Dual Degree (B. Tech/M.Tech) Thesis, Department of Electrical Engineering, IIT Madras*

Articles and Technical Reports

- **A. Bhagoji** and E. Wenger, “Comment for the Federal Trade Commission Regulation Rule on Commercial Surveillance and Data Security”, doi
- L. Song*, V. Sehwag*, **A. Bhagoji*** and P. Mittal, “A Critical Evaluation of Open-world Machine Learning”, arXiv:2007.04391

- A.B. Aloshious, **A. Bhagoji** and P. Sarvepalli, "On the Local Equivalence of 2D Color Codes and Surface Codes with Applications", arXiv:1804.00866
- C. Sitawarin, **A. Bhagoji**, A. Mosenia, M. Chiang and P. Mittal, "DARTS: Deceiving Autonomous Cars with Toxic Signs", arXiv:1802.06430

Teaching & Mentoring

At University of Chicago

- **Mentoring**: Zain Sarwar (Ph.D. Student), Brennan Schaffner (Ph.D. Student),Christian Cianfarani (Ph. D. student), Emily Wenger (Ph. D. student), Huiying Li (Ph. D. Student), Kyle McMillan (Ph. D. Student), Shawn Shan (Ph.D. student), Shinan Liu (Ph.D. Student), Van Tran (Ph. D. student), Wenxin Ding (Ph. D. student), Emilio Andere (B.A., Computer Science), Josephine Passananti (B.A., Computer Science 2022), Roma Bhattacharjee (Lab School), William Zhu (Lab School)

At Princeton University

- **Mentoring**: Ashwinee Panda (Ph. D. Student), Chawin Sitawarin (B.S.E, Electrical Engineering 2019), Chong Xiang (Ph.D. student), Jacob Brown (Masters), Liwei Song (Ph.D. student), Matteo Russo (B.S.E, Computer Science 2020), Vikash Sehwag (Ph.D. student),
- Teaching Assistant for *ELE535: Machine Learning and Pattern Recognition* [Fall 2017]

At the Indian Institute of Technology, Madras

- Teaching Assistant for *EE5121: Convex Optimization* [Spring 2015]
- Teaching Assistant for *EE5701: Advanced Communications Lab* [Fall 2014]

Professional Service & Development

- **Area Chair** for Neural Information Processing Systems 2023
- **Reviewer**: CCS 2024, AAAI 2019, CVPR 2020-21, ECCV 2020, ICCV 2019, IEEE Transactions on Information Forensics, IEEE Transactions on Information Theory, ICML 2020-22, Neurips 2019-22, NeurIPS MLITS Workshop 2018-19, Journal on Machine Learning Research
- Volunteer with Code Your Dreams teaching underprivileged youth data science
- Participated in the Leadership and Management in Action Program (L-MAP) for Postdoctoral Researchers
- Selected for 1st *Ethics of AI* Professional Development Learning Cohort at Princeton University

References[1]

- Prateek Mittal,
  Professor,
  Princeton University
- Ben Zhao,
  Professor,
  University of Chicago
- Nick Feamster,
  Professor,
  University of Chicago
- Supriyo Chakraborty,
  Distinguished Applied Researcher,
  Capital One
- Daniel Cullina,
  Assistant Professor,
  Pennsylvania State University

[1] Contact details available on request