Eric Wong

Phone: +1 (339) 223-7159
Email: wongeric@mit.edu
Website: https://riceric22.github.io
Google Scholar: [link]

Last updated: 2022/07/15

Briefly

How can we make sure that deep learning models are actually doing what we want them to do? My research interests are centered around foundations of reliable machine learning systems: understanding, debugging, and guaranteeing the behavior of data-driven models. I created the first provable defenses that guarantee robustness to adversarial examples and real-world specifications.

Education

University of Pennsylvania, Assistant Professor — Philadelphia, PA, *2022-current*

Massachusetts Institute of Technology, Post-Doctoral Associate — Cambridge, MA, *2020-2022*
Advisor: Aleksander Mądry

Carnegie Mellon University, Ph. D. in Machine Learning — Pittsburgh, PA, *2015-2020*
Thesis: Provable, structured, and efficient methods for robustness of deep networks to adversarial examples; SCS Dissertation Award — Honorable Mention
Advisor: Zico Kolter

Carnegie Mellon University, B. S. in Computer Science — Pittsburgh, PA, *2011-2015*
Double major in Mathematics, minor in Machine Learning

Selected publications

2021 **ICML** **Leveraging sparse linear layers for debuggable deep networks**
Eric Wong, Shibani Santurkar, and Aleksander Madry
International Conference on Machine Learning

ICLR **Learning perturbation sets for robust machine learning**
Eric Wong and J Zico Kolter
International Conference on Learning Representations

2020 **ICML** **Overfitting in adversarially robust deep learning**
Leslie Rice, Eric Wong, and Zico Kolter
International Conference on Machine Learning

ICLR **Fast is better than free: Revisiting adversarial training**
Eric Wong, Leslie Rice, and J. Zico Kolter
International Conference on Learning Representations

2019 **ICML** **Wasserstein adversarial examples via projected sinkhorn iterations**
Eric Wong, Frank Schmidt, and Zico Kolter
International Conference on Machine Learning

2018 **NeurIPS** **Scaling provable adversarial defenses**
Eric Wong, Frank Schmidt, Jan Hendrik Metzen, and J. Zico Kolter
Advances in Neural Information Processing Systems

ICML **Provable defenses against adversarial examples via the convex outer adversarial polytope**
Eric Wong and Zico Kolter
International Conference on Machine Learning

Teaching experience

2016 **Practical Data Science**
Designed new assignments, taught recitations, and prepared write-ups for the first iteration of CMU's Practical Data Science course (15-388/688). I was the head TA (out of two TAs) and managed over 300 enrolled students. Received 55 student reviews with an average rating of 4.85/5.

2016-2019 **Eberly Center**
Enrolled in teaching seminars at the Eberly Center in CMU to develop personal teaching skills; seminars include "Teaching Inclusively: Leveraging Diversity and Promoting Equity in Your Classroom" and "Helping Students Develop Mastery and Critical Thinking"

2015 **Advanced Introduction to Machine Learning**
Taught recitations, held office hours, and created/graded assignments for the second iteration of the Advanced Introduction to Machine Learning course intended for doctoral students in CMU's Machine Learning Department.

2014 **Algorithm Design and Analysis**
Taught recitations, conducted oral examinations/office hours, and graded assignments/exams for the computer science department's Algorithm Design and Analysis course (15-451) at CMU.

2013, 2014 **Pervasive and Mobile Computing Services**
Held office hours and graded assignments/exams for the software engineering department's Pervasive and Mobile Computing Services course (08-766/781) at CMU.

2013 **Mobile Development for iOS and Android**
Held office hours and graded assignments/exams for the software engineering department's Mobile Deveopment course (08-723) at CMU.

Work experience

2019-2020 **Bosch Center for Artifical Intelligence (Renningen, Germany and Pittsburgh, PA)**
Created a virtual sensor based on neural networks for a fuel injection system in truck engines; *formally verified the worst-case error of the system* under conversative estimates of physical sensor noise.

2012-2015 **CERT Program (Pittsburgh, PA)**
Migrated secure coding rules from POSIX to C11; analyzed security reports for Java android applications; developed an analysis toolfor security vulnerabilities in source code.

Community service, outreach, and mentorship

2021-2022 **MIT Undergraduate Research Opportunities Program**
Directly supervised an undergraduate for the UROP program at MIT; provided an opportunity for a member of an under-represented group to learn about machine learning and tackle a challenging research project
https://urop.mit.edu/

2021, 2022 **MEnTorEd Opportunities in Research (METEOR Program, MIT)**
Participated in the METEOR postdoc fellowship selection committee, an on-going effort at CSAIL MIT to increase diversity, equity, and inclusion. Provided confidential technical feedback on candidates based on their application materials.
https://www.csail.mit.edu/meteor

2020-2022 **PhD student mentoring (MIT)**
Mentored and collaborated with three PhD students in various research projects at Aleksander Mądry's lab at MIT; guided them through the research process, paper writing, and publication cycle; has resulted in two submissions that are under review at machine learning conferences.

2021 **Graduate Application Assistance Program**
Assisted applicants from under-represented groups with their graduate student applications to MIT's EECS PhD program.
https://www.thrive-eecs.mit.edu/gaap

2020-2021 **Masters student mentoring (CMU)**
Directly supervised a Masters student in the Robotics Institute at CMU; lead student through near completion of a research project, resulting in a conference paper currently under submission; the student is now a PhD student at the Computer Science Department at Stanford University.

2019-2020 **PhD student mentoring (CMU)**
Mentored and collaborated with a first-year PhD student in the Computer Science Department at CMU; guided the student through the entire publication process, resulting in two conference papers published ICLR and ICML.

2019-2020 **CMU AI Mentoring Program**
Mentored undergraduate women and minorities in one-on-one meetings to provide career advice and discuss research/graduate school; the mentee is now a PhD student at UC Berkeley.

2019-2020 **Teknowledge at Obama Academy**
Taught middle schoolers how to code as part of the Teknowledge outreach program at the Obama Academy; courses were intended to provide early exposure to computer science for under-represented students in low-income neighborhoods of Pittsburgh

2019	**Mental Health First Aid Certification** Underwent training to recognize mental health issues and provide first aid assistance to those in need https://www.mentalhealthfirstaid.org/
2018-2020	**Undergraduate student mentoring (CMU)** Directly supervised a visiting undergraduate from IIT Delhi on a research project at CMU; continued guiding and supporting the student throughout the publication cycle culminating in the completion of the project and an ICML 2020 publication; the student is now a PhD student in the Machine Learning Department at Carnegie Mellon University.

Professional service

2022	**New Frontiers in Adversarial Machine Learning** Organizer for an ICML 2022 workshop on new directions in adversarial machine learning
2022	**Workshop on Adversarial Machine Learning and Beyond** Organizer for an AAAI 2022 workshop broadly themed around adversarial machine learning Website: https://advml-workshop.github.io/aaai2022/
2021	**A Blessing in Disguise: The Prospects and Perils of Adversarial Machine Learning** Organizer for an ICML 2021 workshop themed around the dangers and benefits of adversarial machine learning. Website: https://advml-workshop.github.io/icml2021/
2021	**Robust and reliable ML in the real world** Main organizer for an ICLR 2021 workshop on real world robustness. Website: https://sites.google.com/connect.hku.hk/robustml-2021/home
2022	**Principles of Distribution Shift Workshop at ICML** Program commitee Website: https://sites.google.com/view/icml-2022-pods
2021	**14th ACM workshop on Artificial Intelligence and Security** Program commitee Website: https://aisec.cc/
2020	**Towards Trustworthy ML: Rethinking Security and Privacy for ML at ICLR** Program commitee Website: https://trustworthyiclr20.github.io/
2020-Present	**ICML, NeurIPS, ICLR** Reviewer
2020	**AAAI** Program committee

2019	**Human-Centric Machine Learning Workshop at NeurIPS** Program commitee Website: https://sites.google.com/view/hcml-2019
2019	**Security and Privacy of Machine Learning Workshop at ICML** Program commitee Website: https://icml2019workshop.github.io/
2019	**Adversarial Machine Learning in Real-World Computer Vision Systems at CVPR** Technical program committee
2019	**1st Workshop on Adversarial Learning Methods for Machine Learning and Data Mining at KDD** Technical program committee Website: https://sites.google.com/view/advml
2019	**Safe Machine Learning Workshop at ICLR** Program committee Website: https://sites.google.com/view/safeml-iclr2019/

Awards

2020	**SCS Dissertation Award — Honorable Mention**, *Provable, structured, and efficient methods for robustness of deep networks to adversarial examples*, Carnegie Mellon University
2020	**Siebel Scholar Fellowship**, Carnegie Mellon University
2017	**Best Defense Paper at NeurIPS 2017 ML & Security Workshop**, *Provable defenses against adversarial examples via the convex outer adversarial polytope*
2013	**Summer Undergraduate Reesearch Fellowship**, Carnegie Mellon University

Invited talks

2022	**Invited talk at TrustML Young Scientist Seminar** Title: *Debugging Machine Learning*. Website: https://trustmlresearch.github.io/
2022	**Invited talk at the 56th Annual Conference on Information Sciences and Systems (CISS)** Title: *Robustness for the real world*. Website: https://ee-ciss.princeton.edu/
2021	**Panelist at ATVA 2021 Workshop on Security and Reliability of Machine Learning**. Website: https://sites.google.com/view/srml-atva2021

All publications

2022 **arXiv** **A Data-Based Perspective on Transfer Learning**
Saachi Jain, Hadi Salman, Alaa Khaddaj, Eric Wong, Sung Min Park, and Aleksander Madry
arXiv:2207.02842

arXiv **When does Bias Transfer in Transfer Learning**
Hadi Salman, Saachi Jain, Andrew Ilyas, Logan Engstrom, Eric Wong, and Aleksander Madry
arXiv:2207.02842

IEEE OJCS **DeepSplit: Scalable verification of deep neural networks via operator splitting**
Shaoru Chen, Eric Wong, J Zico Kolter, and Mahyar Fazlyab
IEEE Open Journal of Control Systems

ICLR **Missingness bias in model debugging**
Saachi Jain, Hadi Salman, Eric Wong, and Aleksander Madry

CVPR **Certified patch robustness via smoothed vision transformers**
Hadi Salman, Saachi Jain, Eric Wong, and Aleksander Mądry
arXiv:2110.07719 [cs.CV]

2021 **arXiv** **Adversarial robustness in discontinuous spaces via alternating sampling & descent**
Rahul Venkatesh, Eric Wong, and J Zico Kolter

ICML **Leveraging sparse linear layers for debuggable deep networks**
Eric Wong, Shibani Santurkar, and Aleksander Madry
International Conference on Machine Learning

ICLR **Learning perturbation sets for robust machine learning**
Eric Wong and J Zico Kolter
International Conference on Learning Representations

2020 **ICML** **Adversarial robustness against the union of multiple perturbation models**
Pratyush Maini, Eric Wong, and Zico Kolter
International Conference on Machine Learning

DiffCVGP **Semantic Adversarial Robustness with Differentiable Ray-Tracing**
Rahul Venkatesh, Eric Wong, and J Zico Kolter
Workshop on Differentiable Vision, Graphics, and Physics in Machine Learning at NeurIPS 2020 (Workshop paper)

ICML **Overfitting in adversarially robust deep learning**
Leslie Rice, Eric Wong, and Zico Kolter
International Conference on Machine Learning

ICLR **Fast is better than free: Revisiting adversarial training**
Eric Wong, Leslie Rice, and J. Zico Kolter
International Conference on Learning Representations

	IEEE IV	**Neural network virtual sensors for fuel injection quantities with provable performance specifications** Eric Wong, Tim Schneider, Joerg Schmitt, Frank R Schmidt, and J Zico Kolter *2020 IEEE Intelligent Vehicles Symposium (IV)*
2019	**ICML**	**Wasserstein adversarial examples via projected sinkhorn iterations** Eric Wong, Frank Schmidt, and Zico Kolter *International Conference on Machine Learning*
2018	**NeurIPS**	**Scaling provable adversarial defenses** Eric Wong, Frank Schmidt, Jan Hendrik Metzen, and J. Zico Kolter *Advances in Neural Information Processing Systems*
	ICML	**Provable defenses against adversarial examples via the convex outer adversarial polytope** Eric Wong and Zico Kolter *International Conference on Machine Learning*
2017	**ICML**	**A semismooth Newton method for fast, generic convex programming** Alnur Ali, Eric Wong, and J Zico Kolter *International Conference on Machine Learning*
2015	**AAAI**	**An SVD and derivative kernel approach to learning from geometric data** Eric Wong and J Zico Kolter *Twenty-Ninth AAAI Conference on Artificial Intelligence*