

Huan Zhang

Post-doctoral Fellow
7223 Gates and Hillman Centers
Department of Computer Science, Carnegie Mellon University
Pittsburgh, PA 15213

✉ huanzhan@cs.cmu.edu huan-zhang.com

PROFESSIONAL APPOINTMENTS

Post-Doctoral Fellow at Carnegie Mellon University

Pittsburgh, PA, 2021 - Present.

Supervisor: [J. Zico Kolter](#)

EDUCATION

Ph.D. in Computer Science

UCLA (2020).

Advisor: [Cho-Jui Hsieh](#)

Area: formal verification methods for artificial intelligence (AI); AI safety and security; trustworthy AI.

M.S. in Computer Engineering

UC Davis (2014).

Advisor: [Venkatesh Akella](#)

Area: computer architecture, parallel computing, scalable machine learning.

Bachelor of Engineering

Zhejiang University (2012).

Major: Information Engineering & Optical Engineering

EMPLOYMENT

Internship at Google DeepMind

London, UK, Summer 2019.

Mentor: [Krishnamurthy \(Dj\) Dvijotham](#) and [Po-Sen Huang](#)

Internship at Microsoft Research

Redmond, WA, Summer 2018.

Mentor: [Pengchuan Zhang](#) and [Lin Xiao](#)

Internship at Amazon A9.com

Palo Alto, CA, Spring 2018.

Mentor: [Inderjit Dhillon](#)

Internship at IBM T.J. Watson Research Center

Yorktown Heights, NY, Summer 2017, 2018.

Mentor: [Jinfeng Yi](#) and [Pin-Yu Chen](#)

Internship at Nokia Bell Labs

Murray Hill, NJ, Summer 2013, 2015.

Mentor: [Noriaki Kaneda](#) and [Young-Kai Chen](#)

AWARDS AND HONORS


2022 **Schmidt Futures AI2050 Early Career Fellowship**, with a research grant of **\$300,000**

2022 **First Place**, Third International Verification of Neural Networks Competition ([VNN-COMP 2022](#)), team lead.

2021 **First Place**, Second International Verification of Neural Networks Competition ([VNN-COMP 2021](#)), team lead.

- 2021 **Adversarial Machine Learning (AdvML) Rising Star Award**, sponsored by [MIT-IBM Watson AI Lab](#)
- 2018 **IBM PhD Fellowship**, with a stipend of \$60,000.00
- 2011 **National Merit Scholarship**, Ministry of Education, China, awarded to top 2% students.
- 2010 **Meritorious Winner**, [The U.S. Mathematical Contest in Modeling, 2010](#).
- 2009 **National Merit Scholarship**, Ministry of Education, China, awarded to top 2% students.
- 2009 **First Prize**, China Undergraduate Mathematical Contest in Modeling, 2009.
- 2009 **Second Prize**, East China Undergraduate Mathematical Contest in Modeling, 2009.

PUBLICATIONS

 **Google Scholar Profile**: number of citations **8500+**, h-index **37**, i10-index **44**

Peer-reviewed Conference papers (* indicates **co-first** authors)

- 2022 **General Cutting Planes for Bound-Propagation-Based Neural Network Verification**
Huan Zhang*, Shiqi Wang*, Kaidi Xu*, Linyi Li, Bo Li, Suman Jana, Cho-Jui Hsieh and Zico Kolter
Advances in Neural Information Processing Systems (NeurIPS)
- 2022 **Are AlphaZero-like Agents Robust to Adversarial Perturbations?**
Li-Cheng Lan, Huan Zhang, Ti-Rong Wu, Meng-Yu Tsai, I-Chen Wu, Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)
- 2022 **Efficiently Computing Local Lipschitz Constants of Neural Networks via Bound Propagation**
Zhouxing Shi, Yihan Wang, Huan Zhang, Zico Kolter, Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)
- 2022 **δ -SAM: Sharpness-Aware Minimization with Dynamic Reweighting**
Wenxuan Zhou, Fangyu Liu, Huan Zhang, Muhao Chen
Findings in Empirical Methods in Natural Language Processing (EMNLP)
- 2022 **A Branch and Bound Framework for Stronger Adversarial Attacks of ReLU Networks**
Huan Zhang*, Shiqi Wang*, Kaidi Xu, Yihan Wang, Suman Jana, Cho-Jui Hsieh and Zico Kolter
International Conference on Machine Learning (ICML)
- 2022 **Linearity Grafting: Relaxed Neuron Pruning Helps Certifiable Robustness**
Tianlong Chen*, Huan Zhang*, Zhenyu Zhang, Shiyu Chang, Sijia Liu, Pin-Yu Chen and Zhangyang Wang
International Conference on Machine Learning (ICML)
- 2022 **ViP: Unified Certified Detection and Recovery for Patch Attack with Vision Transformers**
Junbo Li, Huan Zhang, Cihang Xie
European Conference on Computer Vision (ECCV)

- 2022 **COPA: Certifying Robust Policies for Offline Reinforcement Learning against Poisoning Attacks**
Fan Wu, Linyi Li, Chejian Xu, Huan Zhang, Bhavya Kailkhura, Krishnaram Kenthapadi, Ding Zhao and Bo Li
International Conference on Learning Representations (ICLR)
- 2021 **Beta-CROWN: Efficient Bound Propagation with Per-neuron Split Constraints for Complete and Incomplete Neural Network Verification**
Shiqi Wang*, Huan Zhang*, Kaidi Xu*, Xue Lin, Suman Jana, Cho-Jui Hsieh and Zico Kolter
Advances in Neural Information Processing Systems (NeurIPS)
- 2021 **Training Certifiably Robust Neural Networks with Efficient Local Lipschitz Bounds**
Yujia Huang, Huan Zhang, Yuanyuan Shi, Zico Kolter and Anima Anandkumar
Advances in Neural Information Processing Systems (NeurIPS)
- 2021 **Robustness Between the Worst and Average Case**
Leslie Rice, Anna Bair, Huan Zhang, and Zico Kolter
Advances in Neural Information Processing Systems (NeurIPS)
- 2021 **Fast Certified Robust Training via Better Initialization and Shorter Warmup**
Zhouxing Shi*, Yihan Wang*, Huan Zhang, Jinfeng Yi and Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)
- 2021 **Robust Reinforcement Learning on State Observations with Learned Optimal Adversary**
Huan Zhang*, Hongge Chen*, Duane Boning, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2021 **Fast and Complete: Enabling Complete Neural Network Verification with Rapid and Massively Parallel Incomplete Verifiers**
Kaidi Xu*, Huan Zhang*, Shiqi Wang, Yihan Wang, Suman Jana, Xue Lin, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2021 **Double Perturbation: On the Robustness of Robustness and Counterfactual Bias Evaluation**
Chong Zhang, Jieyu Zhao, Huan Zhang, Kai-Wei Chang, Cho-Jui Hsieh
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)
- 2020 **Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations**
Huan Zhang*, Hongge Chen*, Chaowei Xiao, Bo Li, Duane Boning, Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)
- 2020 **Automatic Perturbation Analysis for Scalable Certified Robustness and Beyond**
Kaidi Xu*, Zhouxing Shi*, Huan Zhang*, Yihan Wang, Minlie Huang, Kai-Wei Chang, Bhavya Kailkhura, Xue Lin, Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)
- 2020 **An Efficient Adversarial Attack for Tree Ensembles**
Chong Zhang, Huan Zhang, Cho-Jui Hsieh
Advances in Neural Information Processing Systems (NeurIPS)

- 2020 **Reducing Sentiment Bias in Language Models via Counterfactual Evaluation**
Po-Sen Huang*, Huan Zhang*, Ray Jiang, Robert Stanforth, Johannes Welbl, Jack Rae, Vishal Maini, Dani Yogatama, Pushmeet Kohli
Findings in Empirical Methods in Natural Language Processing (EMNLP)
- 2020 **On ℓ_p -norm Robustness of Ensemble Decision Stumps and Trees**
Yihan Wang, Huan Zhang, Hongge Chen, Duane Boning and Cho-Jui Hsieh
International Conference on Machine Learning (ICML)
- 2020 **Towards Stable and Efficient Training of Verifiably Robust Neural Networks**
Huan Zhang, Hongge Chen, Chaowei Xiao, Sven Gowal, Robert Stanforth, Bo Li, Duane Boning, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2020 **Robustness Verification for Transformers**
Zhouxing Shi, Huan Zhang, Kai-Wei Chang, Minlie Huang, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2020 **MACER: Attack-free and Scalable Robust Training via Maximizing Certified Radius**
Runtian Zhai, Chen Dan, Di He, Huan Zhang, Boqing Gong, Pradeep Ravikumar, Cho-Jui Hsieh, Liwei Wang
International Conference on Learning Representations (ICLR)
- 2019 **Robustness Verification of Tree-based Models**
Hongge Chen*, Huan Zhang*, Si Si, Yang Li, Duane Boning, Cho-Jui Hsieh.
Advances in Neural Information Processing Systems (NeurIPS)
- 2019 **A Convex Relaxation Barrier to Tight Robustness Verification of Neural Networks**
Hadi Salman, Greg Yang, Huan Zhang, Cho-Jui Hsieh, Pengchuan Zhang
Advances in Neural Information Processing Systems (NeurIPS)
- 2019 **Provably Robust Deep Learning via Adversarially Trained Smoothed Classifiers**
Hadi Salman, Greg Yang, Jerry Li, Pengchuan Zhang, Huan Zhang, Ilya Razenshteyn, Sebastien Bubeck
Advances in Neural Information Processing Systems (NeurIPS)
- 2019 **The Limitations of Adversarial Training and the Blind-Spot Attack**
Huan Zhang*, Hongge Chen*, Zhao Song, Duane Boning, Inderjit Dhillon, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2019 **Query-Efficient Hard-label Black-box Attack: An Optimization-based Approach**
Minhao Cheng, Thong Le, Pin-Yu Chen, Huan Zhang, Jinfeng Yi, Cho-Jui Hsieh
International Conference on Learning Representations (ICLR)
- 2019 **Structured Adversarial Attack: Towards General Implementation and Better Interpretability**
Kaidi Xu, Sijia Liu, Pu Zhao, Pin-Yu Chen, Huan Zhang, Quanfu Fan, Deniz Erdogmus, Yanzhi Wang, Xue Lin
International Conference on Learning Representations (ICLR)

- 2019 **Robust Decision Trees Against Adversarial Examples**
Hongge Chen, Huan Zhang, Duane Boning, Cho-Jui Hsieh
International Conference on Machine Learning (ICML)
- 2019 **Evaluating Robustness of Deep Image Super-Resolution Against Adversarial Attacks**
Jun-Ho Choi, Huan Zhang, Jun-Hyuk Kim, Cho-Jui Hsieh, Jong-Seok Lee
International Conference on Computer Vision (ICCV)
- 2019 **Second Rethinking of Network Pruning in the Adversarial Setting**
Shaokai Ye, Kaidi Xu, Sijia Liu, Hao Cheng, Jan-Henrik Lambrechts, Huan Zhang, Aojun Zhou, Kaisheng Ma, Yanzhi Wang, Xue Lin
International Conference on Computer Vision (ICCV)
- 2019 **RecurJac: An Efficient Recursive Algorithm for Bounding Jacobian Matrix of Neural Networks and Its Applications**
Huan Zhang, Pengchuan Zhang, Cho-Jui Hsieh
AAAI Conference on Artificial Intelligence (AAAI)
- 2019 **AutoZOOM: Autoencoder-based Zeroth Order Optimization Method for Attacking Black-box Neural Networks**
Chun-Chen Tu, Paishun Ting, Pin-Yu Chen, Sijia Liu, Huan Zhang, Jinfeng Yi, Cho-Jui Hsieh, Shin-Ming Cheng
AAAI Conference on Artificial Intelligence (AAAI)
- 2018 **Efficient Neural Network Robustness Certification with General Activation Functions**
Huan Zhang*, Tsui-Wei Weng*, Pin-Yu Chen, Cho-Jui Hsieh, Luca Daniel.
Advances in Neural Information Processing Systems (NIPS)
- 2018 **Towards Fast Computation of Certified Robustness for ReLU Networks**
Tsui-Wei Weng*, Huan Zhang*, Hongge Chen, Zhao Song, Cho-Jui Hsieh, Duane Boning, Inderjit S. Dhillon, Luca Daniel.
International Conference on Machine Learning (ICML)
- 2018 **Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach**
Tsui-Wei Weng*, Huan Zhang*, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, Luca Daniel
International Conference on Learning Representations (ICLR)
- 2018 **Attacking Visual Language Grounding with Adversarial Examples: A Case Study on Neural Image Captioning**
Hongge Chen*, Huan Zhang*, Pin-Yu Chen, Jinfeng Yi, Cho-Jui Hsieh
56th Annual Meeting of the Association for Computational Linguistics (ACL)
- 2018 **Is Robustness the Cost of Accuracy? Lessons Learned from 18 Deep Image Classifiers**
Dong Su*, Huan Zhang*, Hongge Chen, Jinfeng Yi, Pin-Yu Chen, Yupeng Gao.
European Conference on Computer Vision (ECCV)
- 2018 **Towards Robust Neural Networks via Random Self-ensemble**
Xuanqing Liu, Minhao Cheng, Huan Zhang, Cho-Jui Hsieh
European Conference on Computer Vision (ECCV)

- 2018 **EAD: Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples**
Pin-Yu Chen, Yash Sharma, Huan Zhang, Jinfeng Yi and Cho-Jui Hsieh
In AAAI Conference on Artificial Intelligence (AAAI)
- 2018 **GPU-acceleration for Large-scale Tree Boosting**
Huan Zhang, Si Si and Cho-Jui Hsieh
SysML Conference
- 2017 **Gradient Boosted Decision Trees for High Dimensional Sparse Output**
Si Si, Huan Zhang, Sathiya Keerthi, Dhruv Mahajan, Inderjit Dhillon and Cho-Jui Hsieh
34th International Conference on Machine Learning (ICML)
- 2017 **Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent**
Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang and Ji Liu
Advances in Neural Information Processing Systems (NIPS)
- 2016 **HogWild++: A New Mechanism for Decentralized Asynchronous Stochastic Gradient Descent**
Huan Zhang, Cho-Jui Hsieh, Venkatesh Akella
IEEE International Conference on Data Mining (ICDM)
- 2016 **Fixing the Convergence Problems in Parallel Asynchronous Dual Coordinate Descent**
Huan Zhang, Cho-Jui Hsieh
IEEE International Conference on Data Mining (ICDM)
- 2016 **Sublinear Time Orthogonal Tensor Decomposition**
Zhao Song, David P. Woodruff, Huan Zhang
Advances in Neural Information Processing Systems (NIPS)
- 2016 **A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order**
Xiangru Lian, Huan Zhang, Cho-Jui Hsieh, Yijun Huang and Ji Liu
Advances in Neural Information Processing Systems (NIPS)

Peer-reviewed workshop papers (* indicates co-first authors)

- 2019 **Enhancing Certifiable Robustness via a Deep Model Ensemble**
Huan Zhang, Minhao Cheng and Cho-Jui Hsieh
ICLR 2019 Safe Machine Learning Workshop
- 2018 **Realtime Query Completion via Deep Language Models**
Po-Wei Wang, Huan Zhang, Vijai Mohan, Inderjit S. Dhillon and J. Zico Kolter
SIGIR Workshop On eCommerce
- 2017 **ZOO: Zeroth Order Optimization based Black-box Attacks to Deep Neural Networks without Training Substitute Models**
Pin-Yu Chen*, Huan Zhang*, Yash Sharma, Jinfeng Yi and Cho-Jui Hsieh
10th ACM Workshop on Artificial Intelligence and Security (Best Paper Finalist)

- 2014 **Burst Mode Processing: An Architectural Framework for Improving Performance in Future Chip Microprocessors**
Huan Zhang, Rajeevan Amirtharajah, Christopher Nitta, Matthew Farrens and Venkatesh Akella
Workshop on Managing Overprovisioned Systems, Co-located with ASPLOS-19
- 2013 **HySIM: Towards a Scalable, Accurate and Fast Simulator for Manycore Processors**
Kramer Straube, Huan Zhang, Christopher Nitta, Matthew Farrens and Venkatesh Akella
3rd Workshop on the Intersections of Computer Architecture and Reconfigurable Logic, Co-located with MICRO-46

SELECTED TALKS

- 2022 **Caltech**, DOLCIT Seminar Series, title “Formal Verification of Deep Neural Networks: Challenges and Recent Advances”.
- 2022 **Princeton University**, Virtual Seminars on Security and Privacy in Machine Learning, title “Formal Verification of Deep Neural Networks: Challenges and Recent Advances”.
- 2022 **Johns Hopkins University**, Institute for Assured Autonomy Seminar Series, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2022 **Carnegie Mellon University (CMU)**, AI Seminar, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2021 **University of California Santa Barbara (UCSB)**, Computer Science Colloquium, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2021 **Northeastern University**, Security Seminar, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2021 **University of Illinois at Urbana-Champaign (UIUC)**, Computer Science Speakers Series, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2021 **University of Southern California (USC)**, AI Seminar, title “How Can We Trust a Black-box? A Quest for Scalable and Powerful Neural Network Verifiers”.
- 2021 **Lorentz Center Workshop on Robust Artificial Intelligence**, title “Robust Reinforcement Learning Against Adversarial Perturbations on State Observations”.
- 2021 **Bosch Center for Artificial Intelligence (BCAI)**, title “Complete and Incomplete Neural Network Verification with Efficient Bound Propagations”.
- 2020 **3rd Workshop on Formal Methods for ML-Enabled Autonomous Systems**, title “Robustness Verification for Ensemble Stumps and Trees”.

PROFESSIONAL SERVICES

Workshop Organization

- 2022 **Lead organizer**, *Trustworthy and Socially Responsible Machine Learning (TSRML)*, co-located with NeurIPS 2022.
- 2022 **Lead organizer**, *1st Workshop on Formal Verification of Machine Learning*, co-located with ICML 2022.
- 2022 **Lead organizer**, *Queer in AI Workshop*, co-located with ICML 2022.

- 2022 **Co-organizer**, [Workshop on Socially Responsible Machine Learning](#), co-located with ICLR 2022.
- 2021 **Co-organizer**, [Workshop on Security and Reliability of Machine Learning](#), co-located with 19th International Symposium on Automated Technology for Verification and Analysis (ATVA 2021).

[Conference/Journal Reviewing and Journal Editing](#)

- 2022 **Guest Journal Editor**, Special Issue “Black-Box Algorithms and Their Applications”, MDPI Algorithms.
- 2021 **Guest Journal Editor**, Trustworthy Machine Learning Research Topic, Frontiers in Big Data, 2021.
- 2021 **Senior Program Committee/Area Chair**, AAAI.
- Conference Paper Reviewer/Program Committee**, NIPS 2016, 2018, 2019, 2020, 2021, 2022; ICML 2019, 2020, 2021, 2022; ICLR 2019, 2020, 2021, 2022, 2023; AAAI 2020, 2021, 2022; UAI 2020, 2021; AISTATS 2021, 2022; CVPR 2020, 2021. USENIX 2020.
- Journal Reviewer**, Journal of Machine Learning Research (JMLR), IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), Springer Journal of Machine Learning.

TEACHING EXPERIENCE

[Tutorials](#)

- 2022 **Formal Verification of Deep Neural Networks: Theory and Practice**, AAAI 2022, Tutorial is publicly available at neural-network-verification.com.
- 2021 **auto_LiRPA: An Automatic Neural Network Verification Library**, [Lorentz Center Workshop on Robust Artificial Intelligence](#).

[Guest Lectures](#)

- 2022 **UIUC**, title “Formal Verification of Deep Neural Networks: Challenges and Recent Advances”, for [CS 562: Advanced Topics in Security, Privacy and Machine Learning](#).
- 2020 **Stony Brook University**, title “Complete and Incomplete Neural Network Verification with Efficient Bound Propagations”, for [CSE 510 - Hybrid Systems - Spring 2021](#).
- 2020 **University of Nebraska Lincoln**, title “CROWN: A Linear Relaxation Framework for Neural Network Verification”, for [CSCE 990: Deep Learning and Assured Autonomy Analysis](#).

[Teaching Assistantship](#)

- 2017 **Big Data & High Performance Statistical Computing**, STA 141C, Instructor: [Cho-Jui Hsieh](#).
- 2015 **Probability and Statistical Modeling for Computer Science**, ECS 132, Instructor: [Dipak Ghosal](#).
- 2013 **Parallel Computer Architecture**, EEC 171, Instructor: [John Owens](#).

OPEN SOURCE PROJECTS

α, β -CROWN: A Neural Network Verification Toolbox (2021-) <http://abcrown.org>.

I lead the development of α, β -CROWN, an efficient and scalable neural network verification toolbox that won the highest total score in 2nd and 3rd International Verification of Neural Network Competition (VNN-COMP 2021 and 2022).

AutoLiRPA: A Neural Network Perturbation Analysis Library (2020-) <http://PaperCode.cc/AutoLiRPA>.

I lead the development of AutoLiRPA, an easy-to-use library capable of automatically giving provable bounds under input or weight perturbations for complex neural networks and other general computational functions.

LightGBM on GPU (2016-2017) <https://github.com/huanzhang12/lightgbm-gpu>.

LightGBM is a popular tree boosting package with high efficiency on large-scale datasets. I accelerated its decision tree construction process on GPUs with 7 to 8 times speedup. My code reaches production quality and has been merged into the LightGBM official repository.

REFERENCES

J. Zico Kolter

Associate Professor of Computer Science
Carnegie Mellon University
Contact: zkolter@cs.cmu.edu

Cho-Jui Hsieh

Associate Professor of Computer Science
University of California, Los Angeles
Contact: chohsieh@cs.ucla.edu

Suman Jana

Associate Professor of Computer Science
Columbia University
Contact: suman@cs.columbia.edu

Bo Li

Assistant Professor of Computer Science
University of Illinois Urbana-Champaign
Contact: lbo@illinois.edu

Pin-Yu Chen

Principal Research Scientist
IBM Research
Contact: pin-yu.chen@ibm.com