Junyuan Hong

University of Texas at Austin

+1 (517) 668-5790 jyhong@utexas.edu https://jyhong.gitlab.io

I am a joint postdoctoral fellow in the Institute for Foundations of Machine Learning (IFML) and Wireless Networking and Communications Group (WNCG) at UT Austin. I was selected as one of the MLSys Rising Stars in 2024, and Best Paper Nomination of VLBD 2024. Part of my work in AI for Health is funded by OpenAI Research Access Program.

Education

Ph.D. in Computer Science and Engineering 2023
Michigan State University, East Lansing, MI, USA
Supervisor: Dr. Jiayu Zhou
Committee: Dr. Anil K. Jain, Dr. Sijia Liu, Dr. Zhangyang Wang, Dr. Jiayu Zhou
Thesis: Data-Centric Privacy-Preserving Machine Learning

M.S. in Computer Science 2018
University of Science and Technology of China, Hefei, China
Supervisor: Dr. Huanhuan Chen

B.S. in Physics with minor in Computer Science 2015
University of Science and Technology of China, Hefei, China

Research Interests

AI Safety & Security: Algorithms, analysis, and benchmarks of the trustworthiness of AI/ML systems including privacy, security, robustness, and efficiency.

Human-centric AI for Healthcare: Trustworthy and affordable medical predictive modeling and intervention for cognition dementia diseases, using generative artificial intelligence.

Research Experiences

Aug '23 - Present **Postdoctoral Fellow**, IFML&WNCG at University of Texas, Austin, TX, USA
Supervisor: Dr. Zhangyang (Atlas) Wang
(1) Generative AI for healthcare (dementia diagnosis and prevention);
(2) Democratizing large vision/language models with privacy protection;
(3) Benchmarking trustworthiness of generative AI.

Aug '18 - July'23 **Research Assistant**, Michigan State University, MI, USA
Supervisor: Dr. Jiayu Zhou
(1) Privacy-preserving and fair machine learning systems for heterogeneous healthcare;
(2) Trustworthy machine learning that are affordable, adaptive and beneficial for more, distributed and diverse users.

Feb '22 - Aug '22 **Research Intern**, Sony AI, NY, USA
Mentor: Dr. Lingjuan Lyu
(1) Privacy-preserving cloud training algoirthms that require low computation costs and low privacy risks for edge devices;
(2) Memory-efficient model adaptation algorithms for dynamically-changing test-time environments, which can fit into edge devices.

Aug '15 - Jun '18 **Research Assistant**, University of Science and Technology of China, Hefei, China
Supervisor: Huanhuan Chen
(1) Hardware and software prototype for detecting underground cables;
(2) Implicit data-augmentation in subspaces for human action recognition.

Honors & Awards

2024	Best Paper Nomination, VLDB 2024
	ML and Systems Rising Stars, ML Commons
2023	Top Reviewer, NeurIPS
	Research Enhancement Award, Michigan State University
	The 3rd place in the U.S.-U.K. Privacy-Enhancing Technologies (PETs) prize challenge
	Dissertation Completion Fellowship, Michigan State University
2021	Carl V. Page Memorial Graduate Fellowship, Michigan State University
2015	Outstanding Freshman Scholarship, University of Science and Technology of China

Selected Publications

(† indicates the equal contribution, * indicates (co-)advised students.)

(1) AI/ML for Health & Finance

[LLM Agents'24] **Junyuan Hong**†, Wenqing Zheng†, Han Meng, Siqi Liang, Anqing Chen, Hiroko Dodge, Jiayu Zhou, Zhangyang Wang. A-CONECT: Designing AI-based Conversational Chatbot for Early Dementia Intervention. *ICLR 2024 Workshop on LLM Agents.*
Application: Chatbot designs with virtual humans for early dementia intervention.

[FL4DM'23] Haobo Zhang*, **Junyuan Hong**, Fan Dong, Steve Drew, Liangjie Xue, Jiayu Zhou. A Privacy-Preserving Hybrid Federated Learning Framework for Financial Crime Detection. *International Workshop on Federated Learning for Distributed Data Mining.*
Application: Financial crime analysis from multiple heterogeneous parties.

[KDD'21] **Junyuan Hong**, Zhuangdi Zhu, Shuyang Yu*, Zhangyang Wang, Hiroko Dodge, and Jiayu Zhou. Federated Adversarial Debiasing for Fair and Transferable Representations. *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
Application: Fair MCI disease modeling using distributed sensors on old adults.

[AD'20] **Junyuan Hong**, Jeffrey Kaye, Hiroko H Dodge, Jiayu Zhou. Detecting MCI Using Real-time, Ecologically Valid Data Capture Methodology: How to Improve Scientific Rigor in Digital Biomarker Analysis. *Alzheimer's & Dementia*
Application: MCI disease modeling using sensors on old adults.

(2) Benchmarking AI/ML Safety and Efficiency

[ICML'24] **Junyuan Hong**[†], Jinhao Duan[†], Chenhui Zhang[†], Zhangheng Li[†], Chulin Xie, Kelsey Lieberman, James Diffenderfer, Brian Bartoldson, Ajay Jaiswal, Kaidi Xu, Bhavya Kailkhura, Dan Hendrycks, Dawn Song, Zhangyang Wang, Bo Li. Decoding Compressed Trust: Scrutinizing the Trustworthiness of Efficient LLMs Under Compression. *The Forty-first International Conference on Machine Learning.*
#Safety #LLM

[VLDB'24] Qinbin Li[†], **Junyuan Hong**[†], Chulin Xie[†], Jeffrey Tan[*], Rachel Xin, Junyi Hou, Xavier Yin, Zhun Wang, Dan Hendrycks, Zhangyang Wang, Bo Li, Bingsheng He, Dawn Song. LLM-PBE: Assessing Data Privacy in Large Language Models. *International Conference on Very Large Data Bases.* (**Best Paper Nomination**)
#Privacy #LLM

[ICML'24] Yihua Zhang[†], Pingzhi Li[†], **Junyuan Hong**[†], Jiaxiang Li, Yimeng Zhang, Wenqing Zheng, Pin-Yu Chen, Jason D. Lee, Wotao Yin, Mingyi Hong, Zhangyang Wang, Sijia Liu, Tianlong Chen. Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark. *The Forty-first International Conference on Machine Learning.*
#Memory-Efficient Optimization #LLM

[FL4DM'23] Siqi Liang[*], Jintao Huang, **Junyuan Hong**, Dun Zeng, Jiayu Zhou, Zenglin Xu. FedNoisy: A Federated Noisy Label Learning Benchmark. *International Workshop on Federated Learning for Distributed Data Mining.*
#Federated-Learning #Robustness

(3) Privacy in AI/ML

[ICLR'24] **Junyuan Hong**, Jiachen T. Wang, Chenhui Zhang, Zhangheng Li, Bo Li, Zhangyang Wang. DP-OPT: Make Large Language Model Your Privacy-Preserving Prompt Engineer. *Proceedings of the Twelfth International Conference on Learning Representations* (**Spotlight, top-5%**).
#LLM #Edge-Computing

[SaTML'24] Zhangheng Li[*], **Junyuan Hong**, Bo Li, Zhangyang Wang. Shake to Leak: Amplifying the Generative Privacy Risk through Fine-Tuning. *The 2nd IEEE Conference on Secure and Trustworthy Machine Learning.*
#Generative-AI

[NeurIPS'23] Haobo Zhang[*†], **Junyuan Hong**[†], Yuyang Deng, Mehrdad Mahdavi, Jiayu Zhou. Understanding Deep Gradient Leakage via Inversion Influence Functions. *Proceedings of the Thirty-Seventh Conference on Neural Information Processing Systems.*
#Analysis-Tool

[FAccT'22] **Junyuan Hong**, Zhangyang Wang, and Jiayu Zhou. Dynamic Privacy Budget Allocation Improves Data Efficiency of Differentially Private Gradient Descent. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.*
#Data-Efficiency

[NeurIPS'22] **Junyuan Hong**, Lingjuan Lyu, and Jiayu Zhou, Micheal Spranger. Outsourcing Training without Uploading Data via Efficient Collaborative Open-Source Sampling. *Proceedings of the Thirty-Sixth Conference on Neural Information Processing Systems.*
#Edge-Computing

(4) AI/ML Safety & Security

[Preprint'24] Zhen Xiang, Linzhi Zheng, Yanjie Li, **Junyuan Hong**, Qinbin Li, Han Xie, Jiawei Zhang, Zidi Xiong, Chulin Xie, Carl Yang, Dawn Song., Bo Li. GuardAgent: Safeguard LLM Agents by a Guard Agent via Knowledge-Enabled Reasoning. *Preprint*.
#Safety #LLM-Agent

[TMLR'23] Haotao Wang, **Junyuan Hong**, Jiayu Zhou, and Zhangyang Wang. How Robust is Your Fairness? Evaluating and Sustaining Fairness under Unseen Distribution Shifts. *Transactions on Machine Learning Research*.
#Federated-Learning #Robustness #Fairness

[ICML'23] **Junyuan Hong**[†], Yi Zeng[†], Shuyang Yu[†*], Lingjuan Lyu, Ruoxi Jia, Jiayu Zhou. Revisiting Data-Free Knowledge Distillation with Poisoned Teachers. *Proceedings of Fortieth International Conference on Machine Learning*.
#AI-Security

[AAAI'23] **Junyuan Hong**, Haotao Wang, Zhangyang Wang, and Jiayu Zhou. Federated Robustness Propagation: Sharing Adversarial Robustness in Heterogeneous Federated Learning. *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*.
#Federated-Learning #Robustness #Efficiency

[ICLR'24] Shuyang Yu[*], **Junyuan Hong**, Haobo Zhang, Haotao Wang, Zhangyang Wang, Jiayu Zhou. Safe and Robust Watermark Injection with a Single OoD Image. *Proceedings of the Twelfth International Conference on Learning Representations*.
#Copyright

[ICLR'23] Shuyang Yu[*], **Junyuan Hong**, Haotao Wang, Zhangyang Wang and Jiayu Zhou. Turning the Curse of Heterogeneity in Federated Learning into a Blessing for Out-of-Distribution Detection. *Proceedings of the Eleventh International Conference on Learning Representations* (**Spotlight, top-5%**).
#Federated-Learning #Robustness

[NeurIPS'22] Haotao Wang, **Junyuan Hong**, Aston Zhang, Jiayu Zhou and Zhangyang Wang. Trap and Replace: Defending Backdoor Attacks by Trapping Them into an Easy-to-Replace Subnetwork. *Proceedings of the Thirty-Sixth Conference on Neural Information Processing Systems*.
#AI-Security

Teaching Experiences

Fall 2023	Mentor, "Trustworthy Large Language Models" (undergraduate level), Directed Reading Program (DiRP), University of Texas at Austin Mentoring students on reading papers and condut projects
Spring 2023	Lectures on privacy and federated learning at "CSE847: Machine Learning" (graduate level), Michigan State University
Spring 2021	Teaching Assistant, "CSE847: Machine Learning" (graduate level), Michigan State University Lectures on privacy and federated learning
Fall 2020	Teaching Assistant, "CSE404: Introduction to Machine Learning" (undergraduate level), Michigan State University

Invited Talks & Presentations

2024 (Guest Lecture) Building Conversational AI for Affordable and Accessible Early Dementia Intervention. *AI Health* Course, The School of Information, UT Austin, April, 2024

(Invited Talk) Shake to Leak: Amplifying the Generative Privacy Risk through Fine-Tuning. *Good Systems Symposium: Shaping the Future of Ethical AI*, UT Austin, March, 2024

2023 (Guest Lecture) Foundation Models Meet Data Privacy: Risks and Countermeasures. *Trustworthy Machine Learning* Course, Virginia Tech, November, 2023

(Invited Talk) Backdoor Meets Data-Free Learning. *TMLR Group*, Hong Kong Baptist University, September, 2023

(Invited Talk) MECTA: Memory-Economic Continual Test-Time Model Adaptation. *Computer Vision Talks*, April, 2023

(Oral) Federated Robustness Propagation: Sharing Adversarial Robustness in Heterogeneous Federated Learning, *The Thirty-Seventh AAAI Conference on Artificial Intelligence* (AAAI 2023), Washington D.C., February 2023.

2022 (Poster) Outsourcing Training without Uploading Data via Efficient Collaborative Open-Source Sampling. *The Thirty-seventh Conference on Neural Information Processing Systems* (NeurIPS 2022), November, 2022.

(Invited Talk) Split-Mix Federated Learning for Model Customization, *TrustML Young Scientist Seminars*, RIKEN, July, 2022

(Poster) Efficient Split-Mix Federated Learning for On-demand and In-situ Model Customization, *The Tenth International Conference on Learning Representations* (ICLR 2022), Virtual, April, 2022.

(Poster) Efficient Split-Mix Federated Learning for On-demand and In-situ Model Customization, *Engineering Graduate Research Symposium*, Michigan State University, April, 2022.

(Invited Talk) Efficient Split-Mix Federated Learning for On-demand and In-situ Model Customization, *Sony AI Journal Club*, Virtual, February, 2022.

(Oral) Dynamic Privacy Budget Allocation Improves Data Efficiency of Differentially Private Gradient Descent, *The 2022 ACM Conference on Fairness, Accountability, and Transparency* (FAccT 2022), Virtual, June 2022.

2021 (Talk) Federated Adversarial Debiasing for Fair and Transferable representations, *CSE Graduate Seminar*, Michigan State University, October, 2021

(Oral) Federated Adversarial Debiasing for Fair and Transferable Representations, *The 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (SIGKDD 2021), Virtual, August 2021.

(Poster) Learning model-based privacy protection under budget constraints, *The Thirty-Fifth AAAI Conference on Artificial Intelligence* (AAAI 2021), Virtual, February 2021.

2020 (Invited Talk) Dynamic Policies on Differential Private Learning, *VITA Seminars*, University of Texas at Austin, Virtual, March 2020.

2018 (Oral) Disturbance Grassmann kernels for subspace-based learning, *The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (SIGKDD 2018), London, August 2018.

Media Coverage

2023 At Summit for Democracy, the United States and the United Kingdom Announce Winners of Challenge to Drive Innovation in Privacy-enhancing Technologies That Reinforce Democratic Values, The White House

Privacy-enhancing Research Earns International Attention, MSU Engineering News

Privacy-Enhancing Research Earns International Attention, MSU Office Of Research And Innovation

Professional Activities

Program Chair:

- Lead Chair at the GenAI for Health: Potential, Trust and Policy Compliance workshop co-located with NeurIPS 2024. (`genai4health.github.io/`).
- Co-Organizer at NeurIPS 2024 LLM Privacy Challenge. (`llm-pc.github.io/`)
- Co-Organizer at The LLM and Agent Safety Competition 2024 (NeurIPS). (`www.llmagentsafetycomp24.com/our-story/`)
- Co-Lead Chair at International Joint Workshop on Federated Learning for Data Mining and Graph Analytics (FedKDD) co-located with ACM SIGKDD 2024. (`fedkdd.github.io`)
- Lead Chair at the 1st International Workshop on Federated Learning for Distributed Data Mining co-located with ACM SIGKDD 2023. (`fl4data-mining.github.io`)

Technical Program Committee Member (or Equivalent Reviewer) for Conferences or Journals:

- Annual Conference on Neural Information Processing Systems (NeurIPS): 2022, 2023, 2024
- International Conference on Learning Representations (ICLR): 2023, 2024
- International Conference on Machine Learning (ICML): 2022, 2023, 2024
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD): 2022, 2023, 2024
- Transactions on Knowledge and Data Engineering (TKDE): 2023
- Journal of Artificial Intelligence Research (JAIR): 2023, 2024
- Transactions on Dependable and Secure Computing (TDSC): 2023, 2024
- ACM Transactions on Computing for Healthcare, 2024
- Neurocomputing: 2021, 2022
- ACM Transactions on Computing for Healthcare (ACM Health): 2024
- ACM Transactions on Knowledge Discovery from Data (TKDD): 2020

- International Conference on Artificial Intelligence and Statistics (AISTATS): 2022, 2023
- International Conference on Web Search and Data Mining (WSDM): 2022
- AAAI Conference on Artificial Intelligence (AAAI): 2021, 2022, 2023
- International Joint Conference on Artificial Intelligence (IJCAI): 2019

Volunteers:

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD): 2018, 2021

Advising Students

Students advised or co-advised with my advisor:

2023 - Now **Zhangheng Li**, Ph.D. student, University of Texas at Austin
Research: Privacy of generative AI.

- (First author) Shake to Leak: Amplifying the Generative Privacy Risk through Fine-tuning. *SaTML'24*
- DP-OPT: Make Large Language Model Your Privacy-Preserving Prompt Engineer. *ICLR'24*
- (Co-First author) Decoding Compressed Trust: Scrutinizing the Trustworthiness of Efficient LLMs Under Compression. *ICML'24*

2023 - Now **Runjin Chen**, Ph.D. student, University of Texas at Austin
Research: AI Safety.

- (First author) Extracting and Understanding Superficial Knowledge in Alignment. *Under Review*
- (Co-First author) Unvailing the Low-Rank Structures of LLM Alignment. *Under Review*

2023 - Now **Gabriel Jacob Perin**, Undergraduate student, University of São Paulo, Brazil
Research: AI Safety and Agent for Paper Review.

- (Co-First author) Unvailing the Low-Rank Structures of LLM Alignment. *Under Review*
- (Co-First author) ReviewAgent: Harnessing Public Opinions for Enhancing GenAI Review Qualities on ML Publications. *Under Review*

2023 - 2024 **Jeffrey Ziwei Tan**, Undergraduate student, University of California, Berkeley
Research: Privacy of Large Language Models.

- LLM-PBE: Assessing Data Privacy in Large Language Models. *VLDB'24*

2020 - 2023 **Shuyang Yu**, Ph.D. student, Michigan State University
Research: Federated learning, security and AI copyright.

- (First author) Safe and Robust Watermark Injection with a Single OoD Image. *ICLR'24*
- (First author) Turning the Curse of Heterogeneity in Federated Learning into a Blessing for Out-of-Distribution Detection. *Featured as spotlight at ICLR'23*

- (First author) Who Leaked the Model? Tracking IP Infringers in Accountable Federated Learning. *NeurIPS'23 Workshop on Regulatable ML.*
- Revisiting Data-Free Knowledge Distillation with Poisoned Teachers. *ICML'23.*
- Federated Adversarial Debiasing for Fair and Transferable Representations. *ACM SIGKDD'21.*

2022 - 2023 **Haobo Zhang**, Ph.D. student, Michigan State University
Research: Privacy-preserving machine learning.

- (First author) Won 3rd place at U.S.-U.K. PETs (Privacy-enhancing technologies) Prize Challenge, 2023.
- (First author) Understanding Deep Gradient Leakage via Inversion Influence Functions, *NeurIPS'23.*
- (First author) Privacy-Preserving Hybrid Federated Learning Framework for Financial Crime Detection. *KDD FL4DataMining Workshop'23.*

2022 - 2023 **Siqi Liang**, Ph.D. student, Michigan State University
Research: Robust federated learning.

- (First author) FedNoisy: A Federated Noisy Label Learning Benchmark. *KDD FL4DataMining Workshop'23.*