

Haoran Wang

Last Updated: Dec 10, 2025

Research Interests

My research involves building trustworthy AI systems, with a specific emphasis on improving both factuality and robustness via test-time algorithms. I am broadly interested in fundamental research and interdisciplinary collaborations motivated by important applications, such as automated fact-checking.

Education

- Present **PhD, Computer Science**, *Emory University*, Atlanta, GA, USA.
Thesis Advisor: [Dr. Kai Shu](#)
- 2021 **MS, Computer Science**, *University of Oregon*, Eugene, Oregon, USA.
Thesis Advisor: [Dr. Thien Huu Nguyen](#)
- 2019 **BS, Computer Science**, *Purdue University*, West Lafayette, IN, USA.

Book Chapters and Journal

- [B2] **Haoran Wang**, Baixiang Huang, Kai Shu. [Automated Fact-Checking](#). Chapter in *Oxford Handbook of Misinformation and Disinformation*, Oxford University Press.
- [B1] Baixiang Huang, **Haoran Wang**, Kai Shu. [Factuality of Large Language Models: An Adversarial Perspective](#). Chapter in *Online Trust and Safety: Tools to Combat Online Harms, Misinformation, and Malicious Content*, Taylor and Francis CRC Press.

Conference

- [C9] Baixiang Huang, Zhen Tan, **Haoran Wang**, Zijie Liu, Dawei Li, Ali Payani, Huan Liu, Tianlong Chen, Kai Shu. [Model Editing as a Double-Edged Sword: Steering Agent Ethical Behavior Toward Beneficence or Harm](#). *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 26')*.
- [C8] Yueqing Liang, Liangwei Yang, Chen Wang, Congying Xia, Rui Meng, Xiong Xiao Xu, **Haoran Wang**, Ali Payani, Kai Shu. [Benchmarking LLMs for Political Science: A United Nations Perspective](#). *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 26')*.
- [C7] **Haoran Wang**, Kai Shu. [Spatial-Aware Visual Program Guided Reasoning for Answering Complex Visual Questions](#). *Findings of the Association for Computational Linguistics: IJCNLP-AAACL 2025*.
- [C6] Aman Rangapur, **Haoran Wang**, Ling Jian, Kai Shu. [Fin-Fact: A Benchmark Dataset for Multimodal Financial Fact Checking and Explanation Generation](#). *Companion Proceedings of the ACM Web Conference 2025 (WWW 25')*.
- [C5] **Haoran Wang**, Aman Rangapur, Xiong Xiao Xu, Yueqing Liang, Haroon Gharwi, Carl Yang, Kai Shu. [Piecing It All Together: Verifying Multi-Hop Multimodal Claims](#). In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 25')*.

- [C4] [Haoran Wang](#), Kai Shu. Trojan Activation Attack: Red-Teaming Large Language Models using Steering Vectors for Safety-Alignment. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM 24')*.
- [C3] Yue Huang, Lichao Sun, [Haoran Wang](#), Siyuan Wu, Qihui Zhang, Chujie Gao, et al. TrustLLM: Trustworthiness in Large Language Models. In *Proceedings of the Forty-first International Conference on Machine Learning (ICML 24')*.
- [C2] [Haoran Wang](#), Kai Shu. Explainable Claim Verification via Knowledge-Grounded Reasoning with Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*.
- [C1] [Haoran Wang](#), Yingdong Dou, Canyu Chen, Lichao Sun, Philip S. Yu, Kai Shu. Attacking Fake News Detectors via Manipulating News Social Engagement. In *Proceedings of the ACM Web Conference 2023 (WWW 23')*.

Preprints

- [P8] Baixiang Huang, Limeng Cui, Jiapeng Liu, [Haoran Wang](#), Zhuiyue Tan, Yutong Chen, Chen Luo, Yi Liu, Kai Shu. Towards Effective Model Editing for LLM Personalization. *Preprint 2026*.
- [P7] [Haoran Wang](#), Maryam Khalid, Qiong Wu, Jian Gao, Cheng Cao. Confidence-Guided Fact-Checking with Large Language Models through Probabilistic Certainty and Consistency. *Preprint 2026*.
- [P6] [Haoran Wang](#), Xiong Xiao Xu, Baixiang Huang, Kai Shu. Privacy-Aware Decoding: Mitigating Privacy Leakage of Large Language Models in Retrieval-Augmented Generation. *Preprint 2025*.
- [P5] Xiong Xiao Xu, [Haoran Wang](#), Yueqing Liang, Philip S. Yu, Yue Zhao, Kai Shu. Can Multimodal LLMs Perform Time Series Anomaly Detection?. *Preprint 2025*.
- [P4] Yue Huang, Chujie Gao, Siyuan Wu, [Haoran Wang](#), Xiangqi Wang, Yujun Zhou, Yanbo Wang, Jiayi Ye, Jiawen Shi, Qihui Zhang, Yuan Li, Han Bao, et al. On the Trustworthiness of Generative Foundation Models: Guideline, Assessment, and Perspective. *Preprint 2025*.
- [P3] [Haoran Wang](#), Kai Shu. Beyond Tokens: A Survey on Decoding Methods for Large Language Models and Large Vision-Language Models. *Preprint 2025*.
- [P2] Aman Rangapur, [Haoran Wang](#), Kai Shu. Investigating Online Financial Misinformation and Its Consequences: A Computational Perspective. *Preprint 2023*.
- [P1] Canyu Chen, [Haoran Wang](#), Matthew Shapiro, Yunyu Xiao, Fei Wang, Kai Shu. Combating Health Misinformation in Social Media: Characterization, Detection, Intervention, and Open Issues. *Preprint 2022*.

Research Experience

- May 2025 – **Applied Scientist Intern**, Amazon AGI, Bellevue, WA.
 - Aug 2025
 - Project: [Confidence-Guided LLM Fact-Checker](#)
 - Mentor: Maryam Khalid, Qiong Wu, Jian Gao
 - Manager: Cheng Cao
- Spring 2025 – **Graduate Research Assistant**, Emory University, Atlanta, GA.
 - Present
 - Advisor: Kai Shu
 - Project: [DHS-CAOE](#), sponsored by DHS.
 - Developed efficient methods to mitigate hallucination in LLMs and LVLMs.
- Fall 2022 – **Graduate Research Assistant**, Illinois Institute of Technology, Chicago, IL.
 - May 2024
 - Advisor: Kai Shu
 - Project: [GUISE](#), sponsored by Charles River Analytics, DARPA.
 - Developed systems to extract information flows on social media using a hierarchical template approach.

Fall 2018 –
Spring 2019

Undergraduate Research Assistant, Purdue University, West Lafayette, IN.

- Advisor: Yung-Hsiang Lu
- Project: **CAM2**, sponsored by NSF.
- Evaluated different solutions to Big Data storage problem of unstructured data.
- Built a distributed database to store images and videos along with their metadata captured by network cameras around the globe.

Fellowships & Awards

2025

🏆 **Research Access Program Award**, OpenAI

Open-source Software

TrustGen (contributor): A modular and extensible toolkit for comprehensive trust evaluation of generative foundation models, (100+ Github ★)

TrustLLM (contributor): Trustworthy LLM Benchmark and Toolkit, (500+ Github ★)

Fin-Fact (contributor): Multimodal Financial Fact-Checking Dataset, (Benchmark dataset for shared task at Financial Misinformation Detection workshop at COLING 2025)

Teaching Experience

CS 585: Natural Language Processing, *October 26, 2023*, Illinois Institute of Technology.
Guest Lecturer

CS 550: Advanced Operating Systems, *Fall 2024*, Illinois Institute of Technology.
Graduate Teaching Assistant

CS 211: Computer Science I, *Spring 2020, Winter 2021, Spring 2021*, University of Oregon.
Graduate Teaching Assistant

CS 212: Computer Science II, *Fall 2020*, University of Oregon.
Graduate Teaching Assistant

Academic Service

Session Chair: CIKM 2024

Reviewer/Program Committee: NeurIPS {2025}, ICML {2025}, AAAI {2024, 2025, 2026}, KDD {2024, 2025}, PAKDD{2025}, ACML {2025}

Student Volunteer: ACM FAccT 2023

External Reviewer: SIGIR {2023, 2024}, WSDM{2023}, ICDM{2023}, PAKDD{2024}, SDM{2025}

Mentoring

Iris Qiao, Undergraduate student at Emory University

Aman Rangapur, IIT MS student → Solutions Architect, Allen Institute for AI (AI2)

Hans Guttormsen, Undergraduate student at Illinois Institute of Technology

Technical Skills

Programming languages: Python, Java, C, C++, C#, JavaScript, SQL, Bash, R, Julia

Deep learning frameworks: PyTorch, Hugging Face Transformers, PyTorch Geometric

HPC: CUDA, OpenMP, MPI

Software: Linux, Git, Google Cloud Computing, L^AT_EX