

Minqian Liu

☎ +1 (540)-998-0624 • ✉ minqianliu@vt.edu • 🎓 Google Scholar

Research Summary

My research focuses on building human-centered and responsible frontier AI systems. Totally, I have 15 peer-reviewed publications, including 9 (co-)first-author papers in top-tier venues such as ICLR, ACL, EMNLP, NAACL, COLM, etc. My representative works study critical safety concerns and capability boundaries in generative models, including conducting red-teaming studies on nuanced risks in LLM interactions (COLM'25), developing domain-generalizable guardrails, and creating evaluation methodologies for foundation models (NAACL'24, EMNLP'24).

Education

Virginia Tech

Ph.D. Student in Computer Science

Blacksburg, USA

Aug. 2021-May 2026 (expected)

- Advisor: Lifu Huang and Dawei Zhou
- Thesis Committee: Sebastian Gehrmann, Naren Ramakrishnan, Chris North, Chang-Tien Lu
- Thesis Topic: Holistic and Generalizable Evaluation of Generative Models

South China University of Technology (SCUT)

B.Eng. in Computer Science and Technology

Guangzhou, China

Sept. 2017-Jun. 2021

- Advisor: Mingkui Tan

University of California, Berkeley

Visiting Student

Berkeley, USA

Jan. 2019-May 2019

Employment

Bloomberg CTO Office

Research Intern

New York, USA

Jun. 2025-Sept. 2025

- Mentor: Mark Dredze and Sebastian Gehrmann
- Topic: Domain Generalizable AI Guardrails with Augmented Policy Training

Microsoft Research

Research Intern

Redmond, USA

May 2024-Aug. 2024

- Manager: Dinei Florencio
- Topic: In-context Learning and Visual Chain-of-Thoughts for Vision-Language Models

Amazon AWS AI

Applied Scientist Intern

New York, USA

May 2023-Aug. 2023

- Manager: Patrick Ng
- Topic: Evaluation for Conversational Agents as Database Interfaces

Selected Paper List (by Research Topics)

The full publication list is at 🎓 Google Scholar. * indicates the authors contribute equally.

Human-Centered and Responsible AI

[1] LLMs Can be a Dangerous Persuader: Empirical Study of Persuasion Safety with Large Language Models

- Minqian Liu, Zhiyang Xu, Sarvech Qadir, Heajun An, Qi Zhang, Pamela Wisniewski, Jin-Hee Cho, Ruoxi Jia, Lifu Huang.
- In COLM 2025.

[2] X-EVAL: Generalizable Multi-aspect Text Evaluation via Augmented Instruction Tuning with Auxiliary Evaluation Aspects

- Minqian Liu, Ying Shen, Zhiyang Xu, Yixin Cao, Eunah Cho, Vaibhav Kumar, Reza Ghanadan, Lifu Huang.
- In NAACL (Main), 2024.
- Won **Third Place** in the Scientific Innovation Competition in Alexa Prize SocialBot Grand Challenge 5.

[3] Domain Generalizable AI Guardrails with Augmented Policy Training

- Minqian Liu, Ioana Baldini, David Rabinowitz, David S Rosenberg, Sebastian Gehrmann, Mark Dredze.
- Under Review.

[4] Modeling Theory-of-Mind Dynamics as Process Rewards for Strategic Policy Optimization

- Minqian Liu, Lifu Huang.
- Work in progress.

[5] StagePilot: A Deep Reinforcement Learning Agent for Stage-Controlled Cybergrooming Simulation

- Heajun An, Qi Zhang, Minqian Liu, Xinyi Zhang, Sang Won Lee, Lifu Huang, Pamela Wisniewski, Jin-Hee Cho.
- Under Review.

[6] KnowledgeBot: Improving Assistive Robot for Task Completion and Live Interaction via Neuro-Symbolic Reasoning

- Minqian Liu, Ying Shen, Barry Menglong Yao, Sijia Wang, Jingyuan Qi, Zhiyang Xu, Lifu Huang.
- In *Alexa Prize Simbot Challenge Proceedings*, 2023. (Ranked 4/10 in Alexa Prize SimBot Challenge.)

Multimodal Understanding & Generation

[1] Modality-Specialized Synergizers for Interleaved Vision-Language Generalists

- Minqian Liu*, Zhiyang Xu*, Ying Shen, Joy Rimchala, Jiaxin Zhang, Qifan Wang, Yu Cheng, Lifu Huang.
- In *ICLR 2025*.

[2] ReFocus: Visual Editing as a Chain of Thought for Structured Image Understanding

- Xingyu Fu, Minqian Liu, Zhengyuan Yang, John Richard Corring, Yijuan Lu, Jianwei Yang, Dan Roth, Dinei Florencio, Cha Zhang
- In *ICML 2025*.

[3] Holistic Evaluation for Interleaved Text-and-Image Generation

- Minqian Liu, Zhiyang Xu, Zihao Lin, Trevor Ashby, Joy Rimchala, Jiaxin Zhang, Lifu Huang.
- In *EMNLP (Main)*, 2024.

[4] MultiScript: Multimodal Script Learning for Supporting Open Domain Everyday Tasks

- Minqian Liu*, Jingyuan Qi*, Ying Shen, Zhiyang Xu, Lifu Huang.
- In *AAAI*, 2024.

[5] The Art of SOCRATIC QUESTIONING: Recursive Thinking with Large Language Models

- Jingyuan Qi*, Zhiyang Xu*, Ying Shen, Minqian Liu, Di Jin, Qifan Wang, Lifu Huang.
- In *EMNLP (Main)*, 2023.

NLP Applications

[1] Decomposing and Quantifying Scientific Novelty in Structured Ideation Space

- Minqian Liu*, Yuexi Shen*, Lifu Huang.
- Work in progress.

[2] MetaScientist: A Human-AI Synergistic Framework for Automated Mechanical Metamaterial Design

- Jingyuan Qi*, Zian Jia*, Minqian Liu, Wangzhi Zhan, Junkai Zhang, Xiaofei Wen, Jingru Gan, Jianpeng Chen, Qin Liu, Mingyu Ma, Bangzheng Li, Haohui Wang, Muhao Chen, Dawei Zhou, Ling Li, Wei Wang, Lifu Huang
- In *NAACL Demo Track 2025*.

[3] Teamwork Is Not Always Good: An Empirical Study of Classifier Drift in Class-incremental Information Extraction

- Minqian Liu, Lifu Huang.
- In *ACL (Findings)*, 2023.

[4] Incremental Prompting: Episodic Memory Prompt for Lifelong Event Detection

- Minqian Liu, Shiyu Chang, Lifu Huang.
- In *COLING*, 2022.

Honors & Awards

- Third Place of the Scientific Innovation Competition in Alexa Prize SocialBot Grand Challenge 5. 2023.
- Ranked 4/10 in Public Benchmark Phase of Alexa Prize SimBot Challenge. 2022.
- Saliai Technology Innovation Scholarship (Top 10% in Computer Science Department), SCUT. 2020.
- Outstanding Undergraduate Overseas Study Scholarship, SCUT. 2019.