

# Xiangming Gu

Homepage: <https://guxm2021.github.io>

Email : [xiangming@u.nus.edu](mailto:xiangming@u.nus.edu)

## EDUCATION

---

### National University of Singapore

Singapore

- *PhD candidate, Integrative Sciences and Engineering Programme*  
*Majored in Computer Science, Advisor: Prof. Ye Wang, GPA: 4.80/5.0*

2021/08 – 2025/12 (Expected)

### Tsinghua University

Beijing, China

- *B.E. in Electronic Engineering, B.S. in Finance, GPA: 3.80/4.0*

2017/08 – 2021/06

## EXPERIENCE

---

### Google Deepmind

London, United Kindom

- *Student Researcher*  
2025/05 – Present
  - **Host:** Dr. Petar Veličković and Dr. Larisa Markeeva.
  - **Main Activity:** Projects on (i) adding vision and multimodal support in penzai, which is a JAX toolkit for mechanistic interpretability. (ii) reasoning/test-time-scaling in LLMs.
- **Sea AI Lab**  
*Research Intern*  
Singapore  
2023/03 – 2025/04
  - **Host:** Dr. Tianyu Pang and Dr. Chao Du.
  - **Main Activity:** Projects on (i) memorization in diffusion models, published as TMLR 2025; (ii) infectious jailbreak on (multimodal) LLMs based multi-agent systems, published as ICML 2024; (iii) understanding attention sink in LLMs, published as ICLR 2025 spotlight; (iv) developing evaluation suite for trustworthy datasets based on lm-evaluation-harness; (v) advancing recipe of reinforcement learning with verifiable reward post-training for LLMs.

## RESEARCH (GOOGLE SCHOLAR)

---

\* denotes equal contribution, †denotes correspondence.

### Understanding Generative Models

1. **Xiangming Gu**, Tianyu Pang†, Chao Du, Qian Liu, Fengzhuo Zhang, Cunxiao Du, Ye Wang†, Min Lin. When Attention Sink Emerges in Language Models: An Empirical View. *International Conference on Learning Representations (ICLR)*, 2025. (**Spotlight**)
2. **Xiangming Gu**, Chao Du†, Tianyu Pang†, Chongxuan Li, Min Lin, Ye Wang†. On Memorization in Diffusion Models. *Transactions on Machine Learning Research (TMLR)*, 2025.
3. Federico Barbero\*†, Álvaro Arroyo\*, **Xiangming Gu**, Christos Perivolaropoulos, Michael Bronstein, Petar Veličković, Razvan Pascanu. Why Do LLMs Attend to the First Token? *Conference on Language Modeling (COLM)*, 2025.

### Advancing Generative Models

1. Tongyao Zhu, Qian Liu†, Haonan Wang, Shiqi Chen, **Xiangming Gu**, Tianyu Pang, Min-Yen Kan. SkyLadder: Better and Faster Pretraining via Context Window Scheduling. *International Conference on Learning Representations Workshop on Open Science for Foundation Models (SCI-FM @ ICLR)*, 2025.

### Safety of Generative Models

1. **Xiangming Gu**\*, Xiaosen Zheng\*, Tianyu Pang\*†, Chao Du, Qian Liu, Ye Wang†, Jing Jiang†, Min Lin. Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast. *International Conference on Machine Learning (ICML)*, 2024.
2. Hongfu Liu†, Hengguan Huang, **Xiangming Gu**, Hao Wang, Ye Wang. On Calibration of LLM-based Guard Models for Reliable Content Moderation. *International Conference on Learning Representations (ICLR)*, 2025.

## Speech and Singing

1. **Xiangming Gu**, Longshen Ou, Wei Zeng, Jianan Zhang, Nicholas Wong, Ye Wang<sup>†</sup>. Automatic Lyric Transcription and Automatic Music Transcription from Multimodal Singing. *ACM Transactions on Multimedia Computing Communications and Applications (TOMM)*, 2024.
2. **Xiangming Gu**, Wei Zeng, Ye Wang<sup>†</sup>. Elucidating Gender Fairness in Singing Voice Transcription. *ACM International Conference on Multimedia (MM)*, 2023.
3. **Xiangming Gu\***, Longshen Ou\*, Danielle Ong, Ye Wang<sup>†</sup>. MM-ALT: A Multimodal Automatic Lyric Transcription System. *ACM International Conference on Multimedia (MM)*, 2022. (**Oral, Top Paper Award**)
4. Longshen Ou\*, **Xiangming Gu\***, Ye Wang<sup>†</sup>. Transfer Learning of wav2vec 2.0 for Automatic Lyric Transcription. *International Society for Music Information Retrieval Conference (ISMIR)*, 2022.
5. Yixin Wang, Wei Wei, **Xiangming Gu**, Xiaohong Guan, Ye Wang<sup>†</sup>. Disentangled Adversarial Domain Adaptation for Phonation Mode Detection in Singing and Speech. *IEEE Transactions on Audio, Speech and Language Processing (TASLP)*, 2023.

## Others

1. Hengguan Huang<sup>†</sup>, **Xiangming Gu**, Hao Wang, Chang Xiao, Hongfu Liu, Ye Wang<sup>†</sup>. Extrapolative Continuous-time Bayesian Neural Network for Predictive Streaming Domain Adaptation. *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
2. Wei Wei\*, Hengguan Huang\*, **Xiangming Gu**, Hao Wang, Ye Wang<sup>†</sup>. Unsupervised Mismatch Localization in Cross-Modal Sequential Data with Application to Mispronunciations Localization. *Transactions on Machine Learning Research (TMLR)*, 2022.
3. Youze Xue, Jiansheng Chen<sup>†</sup>, **Xiangming Gu**, Huimin Ma, Hongbing Ma. Boosting Monocular 3D Human Pose Estimation with Part Aware Attention. *IEEE Transactions on Image Processing (TIP)*, 2022.
4. Boyu Zhang, Penghui Yang, **Xiangming Gu**, Hongen Liao<sup>†</sup>. Laser Endoscopic Manipulator Using Spring-reinforced Multi-DoF Soft Actuator. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, also *IEEE Robotics and Automation Letter (RA-L)*, 2021.

## HONORS AND AWARDS

- 
- Dean's Graduate Research Excellence Award (School of Computing, National University of Singapore) 2024
  - Research Incentive Award (School of Computing, National University of Singapore) 2023
  - Research Achievement Award (School of Computing, National University of Singapore) 2025, 2022
  - MM'22 Top Paper Award (Association for Computing Machinery) 2022
  - MM'22 Student Travel Grant (Association for Computing Machinery) 2022
  - President's Graduate Fellowship (National University of Singapore) 2021
  - Visiting Undergraduate Student Scholarship (Tsinghua University) 2020
  - Tsinghua's Friend – Zheng Geru Scholarship (Tsinghua University) 2018

## INVITED TALKS OR POSTERS

---

- Google Deepmind Team Deep Learning: Agent Frontier, invited talk on “Understanding Attention Sink in (Large) Language Models”. 2025
- ASAP Seminar Series, invited talk on “When Attention Sink Emerges in Language Models: An Empirical View”. 2025
- Singapore Alignment Workshop, poster presentation on “Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast”. 2025
- National University of Singapore Research Week Open House, invited talk on “On the Interpretability and Safety of Generative Models”. 2025
- Global Young Scientists Summit, poster presentation on “Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast”. 2025

## PROFESSIONAL SERVICES

---

- **Conference Reviewer:** NeurIPS 2025/2024, ICML 2025, ICLR 2025, CVPR 2025, ICCV 2025/2023, ECCV 2024, ACL ARR 2025/2024, MM 2025/2024, IJCAI 2024, AISTATS 2025/2021
- **Workshop Reviewer:**
  - ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy
  - NeurIPS 2024 Workshop on Attributing Model Behavior at Scale
  - NeurIPS 2024 Safe Generative AI Workshop
- **Journal Reviewer:** TOMM, TASLP, RA-L

## TEACHING

---

- |   |  |
|---|--|
| • <b>Teaching Assistant</b><br><i>CS4347/CS5647, Sound and Music Computing</i>        | National University of Singapore<br><i>Fall 2024</i>   |
| • <b>Teaching Assistant</b><br><i>CS6212, Topics in Media</i>                         | National University of Singapore<br><i>Spring 2024</i> |
| • <b>Teaching Assistant</b><br><i>CS5242, Neural Networks and Deep Learning</i>       | National University of Singapore<br><i>Spring 2023</i> |
| • <b>Teaching Assistant</b><br><i>CS3244: Machine Learning</i>                        | National University of Singapore<br><i>Fall 2022</i>   |
| • <b>Teaching Assistant</b><br><i>CS4243: Computer Vision and Pattern Recognition</i> | National University of Singapore<br><i>Spring 2022</i> |

## TECHNICAL SKILLS

---

- **Coding:** Python, Matlab, Shell, C/C++, HTML, Verilog, Assembly language,  $\text{\LaTeX}$ , ...
- **Libraries:** PyTorch, JAX, Transformers, VLLM, SpeechBrain, ...