Bingqi Shang

Website | Google Scholar | GitHub | bshang@msu.edu

Research Interests

Trustworthy Machine Learning: Machine Unlearning, Alignment & RLHF, Adversarial Machine Learning

Generative AI: Large-language Model, Multi-modal Language Model, Reasoning, Post-training, On-policy Distillation

Education

Michigan State University (MSU) Sep. 2025 - Present
Ph.D., Computer Science Advisor: Prof. Sijia Liu

Northwestern University (NU) Sep. 2023 - Jun. 2025
M.S., Computer Science Advisors: Prof. Qi Zhu and Prof. Xiao Wang

Tongji University Sep. 2019 - Jun. 2023
B.E., Software Engineering School of Computer Science and Technology

Research Experience

LLM Lineage Tracing via Spectral Fingerprints Nov. 2025 - Mar. 2026
Supervisor: Prof. Sijia Liu (MSU)

- Proposed a data-free method to trace LLM lineage from **weight geometry** alone, without any model inputs.
- Spectral energy separates model families; subspace alignment detects shared bases across **110+** model pairs.
- **Publication**: [1].

Backdooring Machine Unlearning via Attention Sinks Apr. 2025 - Oct. 2025
Supervisor: Prof. Sijia Liu (MSU)

- Exposed a vulnerability in LLM unlearning: attention-sink–guided backdoors can **reactivate** forgotten behaviors.
- Exploring applications in unlearned models where backdoor triggers can selectively recover forgotten knowledge.
- **Publication**: [2].

Privacy-Preserving Tuning for Large Models Dec. 2023 - Mar. 2025
Supervisors: Prof. Qi Zhu (NU), Prof. Xiao Wang (NU)

- Developed Split Adaptation (SA) to ensure **data privacy** during adaptation of pre-trained Vision Transformers (ViTs), utilizing bi-level noise injection for privacy-preserving downstream tasks without data sharing.
- Protected **model privacy** by sharing only a low-bit quantized frontend of the ViT, preventing model leakage and ensuring secure adaptation.
- **Publication**: [3].

Professional Experience

MIT-IBM Watson AI Lab *Cambridge, MA* Jun. 2026 - Aug. 2026
Research Scientist Intern, Mentor: Yang Zhang

SAP *Shanghai, China* Jun. 2022 - Mar. 2023
Cloud Developer Intern, Supervisor: April Qi
Project: Cloud Provider Exporter in Go on Kubernetes for AWS, Azure, and GCP, using Prometheus and Grafana.

Publications

*** indicates equal contribution**

[1] Yiwei Chen*, **Bingqi Shang***, Sijia Liu. **Who Built This Model? Tracing LLM Lineage via Spectral Fingerprints in Weight Space**. *COLM'2026*.

[2] **Bingqi Shang***, Yiwei Chen*, Yihua Zhang, Bingquan Shen, Sijia Liu. **Forgetting to Forget: Attention Sink as A Gateway for Backdooring LLM Unlearning**. *COLM'2026*.

[3] Lixu Wang*, **Bingqi Shang***, Yi Li, Payal Mohapatra, Wei Dong, Xiao Wang, Qi Zhu. **Split Adaptation for Pre-trained Vision Transformers**. *CVPR'2025*.

[4] Yancheng Huang*, Changsheng Wang*, Chongyu Fan, Yicheng Lang, **Bingqi Shang**, Yang Zhang, Mingyi Hong, Qing Qu, Alvaro Velasquez, Sijia Liu. **Subspace Control: Turning Constrained Model Steering into Controllable Spectral Optimization**. *COLM'2026*.

[5] Zijie Liu, Jinhao Duan, **Bingqi Shang**, Xinming An, Sijia Liu, Tianlong Chen. **When Retain Constraints Conflict: Mitigating Forget-Retain Interference in Tabular Data**. *EMNLP'2026 Findings*.

Honors

- Shanghai Outstanding Graduate Award 2023
- Outstanding Undergraduate Dissertation Award of Tongji University 2023
- **National Scholarship** (Top 0.2%, highest undergraduate honor in China) 2020

Services

Workshop Co-organizer: AdvML-Frontiers × CoTMA @ COLM'26

Journal Reviewer: IEEE TSP, IEEE TAES

Conference Reviewer: NeurIPS 2026, CVPR 2026, ICASSP 2026–2027

Conference Volunteer: COLM 2026

Personal Interests

Photography 2019 - Present

PROFESSIONAL SKILLS

Programming Languages: Python, Go, C++, Java, Rust, LaTeX, HTML, CSS, JavaScript
Machine Learning Systems: PyTorch, Transformers, Huggingface, W&B, OpenCV, Scikit-learn

Last updated: October 20, 2026.