# Zhimin Li

Nashville, TN | (801) 558-3593

lzm.zhimin@gmail.com | Website | Linkedin


## ACHIEVEMENTS

**Impact:** Delivered **novel visual data analysis tools** for national labs and universities to interpret neural network models ([1], [2]) and study **complex large-scale data** ([3], [4]). Created data pipelines for TB-scale data in large-scale distributed **neural model training**, reducing data storage by 40x and improving training speed by 3x. Designed novel error-aware neural compression methods that reduce the size of scientific data by up to a factor of 1000. Currently, I am working to improve the **trustworthiness of neural network models**, evaluate the visual interpretability of **LLM**, and develop **AI agents** for large-scale visual data analysis.

**Publications**: 5 journal papers, 3 conference papers, 2 workshop papers

## TECHNICAL SKILLS

- **Programming Languages**: Python, Java, C++, C#, SQL, HTML, CSS, JavaScript
- **Data Analysis**: Visualization, NumPy, Pandas, Scikit-Learn, D3, A/B test
- **Machine Learning**: PyTorch, LLM, VLLM, LangChain

## EDUCATION

- **Ph.D. Computer Science** 2016 - 2024
  *University of Utah, Advisor:Valerio Pascucci*
- **B.Sc. Computer Science and Applied Math** 2012 - 2016
  *University of Utah, Member of Pi Mu Epsilon* ***Dean's List***

## EMPLOYMENT

- **Vanderbilt University** Nashville, United States
  *Research Scholar* *Sep 2024 - Now*
  - Designed a compact MLP for scientific data and **achieved 1,000×** and above data compression with high fidelity and error-aware reconstruction.
  - Developed VLATforge, a scalable visualization-centered VQA benchmark that automatically generates diverse charts for robust and prior-knowledge–independent **LLM evaluation**.

- **Lawrence Livermore National Laboratory** Livermore, United States
  *Research Internship* *2016 - 2023*
  - Developed novel algorithms to model the silent error propagation for understanding computation's resiliency by reducing the required samples **up to 16,000 x**.
  - Developed a pioneering visual analysis system for silent data corruption analysis in the high-performance computation environment, which supported **parallel computation** to instantly interact with 3 TB system logs data.
  - Developer of NLIZE, a pioneering interactive visualization system for natural language inference **model interpretation**.
  - Developer of NDDAV, a **high-dimensional data analysis** system used in the national lab environment for complex high-dimensional data analysis through dimension reduction, clustering, and statistical analysis.

## SELECT PROJECTS

- **Speed Up Large Scale Generative Surrogate Model Training**:
  - Built generative models for large-scale scientific simulation data generation and achieved up to **40× data reduction** on multi-terabyte datasets (e.g.,**3.5 TB**) through lossy compression with minor impact on model accuracy.
  - Re-architected I/O and GPU data-loading pipelines, delivering **3× faster training** and lower storage and data movement overhead in HPC environments.
- **Comparing State-of-Art Robustness Neural Network Model Prediction**:

- Delivered a web-based visual analytics system, which compares **thousands of neural network models** at the same time, and reveals insight differences between the robustness of state-of-the-art neural network models.
- Implemented a **scalable, interactive**, and reproducible experimentation pipeline to compare **network pruning** strategies.
- Introduced geometry-based **robustness** metrics that rapidly surface vulnerable samples, enabling targeted robustness evaluation

## PUBLICATIONS

1. **Zhimin, Li**, Harshitha Menon, Charles Fredrick Jekel, Valerio Pascucci and Peter Lindstrom. *Quantifying the impact of lossy compression on neural generative surrogate modeling*. **The 40th IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2026**.
2. **Zhimin, Li**, Shusen Liu, Xin Yu, Kailkhura Bhavya, Jie Cao, James Daniel Diffenderfer, Peer-Timo Bremer, and Valerio Pascucci. *Understanding robustness lottery: A geometric visual comparative analysis of neural network pruning approaches*. **IEEE Transactions on Visualization and Computer Graphics**, 31:6337–6352, 2025.
3. **Zhimin, Li**, Harshitha Menon, Kathryn Mohror, Shusen Liu, Luanzheng Guo, Peer-Timo Bremer, and Valerio Pascucci. *A visual comparison of silent error propagation*. **IEEE Transactions on Visualization and Computer Graphics**, 30:3268–3282, 2022.
4. **Zhimin, Li**, Harshitha Menon, Kathryn Mohror, Peer-Timo Bremer, Yarden Livant, and Valerio Pascucci. *Understanding a program's resiliency through error propagation.* **In Proceedings of the 26th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming**, **PPoPP** '21, page 362–373, New York, NY, USA, 2021. Association for Computing Machinery.
5. **Zhimin, Li**, Harshitha Menon, Kathryn Mohror, Shusen Liu, Luanzheng Guo, Peer-Timo Bremer, and Valerio Pascucci. *Spotsdc: Revealing the silent data corruption propagation in high-performance computing systems*. **IEEE Transactions on Visualization and Computer Graphics**, 27:3938–3952, 2021.
6. Shusen Liu, **Zhimin, Li**, Tao Li, Vivek Srikumar, Valerio Pascucci, and Peer-Timo Bremer. *Nlize: A perturbation-driven visual interrogation tool for analyzing and interpreting natural language inference models*. **IEEE Transactions on Visualization and Computer Graphics**, 25:651–660, 2019.
7. Xiaohan Wang, **Zhimin Li**, Joshua A. Levine, and Matthew Berger. *Seeing the many: Exploring parameter distributions conditioned on features in surrogates*. IEEE Workshop on Uncertainty Visualization: Unraveling Relationships of Uncertainty, AI, and Decision-Making, 2025.
8. Kuangshi Ai, Haichao Miao, **Zhimin, Li**, Chaoli Wang, and Shusen Liu. *An evaluation-centric paradigm for scientific visualization agents*. 1st Workshop on GenAI, Agents, and the Future of VIS (IEEE VIS 2025).
9. Shusen Liu, Haichao Miao, **Zhimin, Li**, M. Olson, V. Pascucci, and P.-T. Bremer. *Ava: Towards autonomous visualization agents through visual perception-driven decision-making*. **Computer Graphics Forum**, 43:e15093, 2024.
10. Shusen Liu, Tao Li, **Zhimin, Li**, Vivek Srikumar, Valerio Pascucci, and Peer-Timo Bremer. *Visual interrogation of attention-based models for natural language inference and machine comprehension*. EMNLP System Demonstrations, 36–41, 2018.

## PREPRINT

1. **Zhimin, Li**, Haichao Miao, Xinyuan Yan, Valerio Pascucci, Matthew Berger, and Shusen Liu. *See or recall: A sanity check for the role of vision in solving visualization understanding tasks with multimodal LLMs*. Under submission, 2025.
2. **Zhimin, Li**, Haichao Miao, Valerio Pascucci, and Shusen Liu. *Visualization literacy of multimodal large language models: A comparative study*. arXiv:2407.10996, 2024.
3. **Zhimin, Li**, Shusen Liu, Kailkhura Bhavya, Peer-Timo Bremer, and Valerio Pascucci. *Instance-wise linearization of neural networks for model interpretation*. arXiv:2310.16295, 2023.

## AWARDS

- **2025**: IEEE Uncertainty Visualization Best Paper Award
- **2016 - 2017**: University of Utah, Ph.D. Fellowship
- **2015**: KLW Artificial/Machine Learning scholarship
- **2015**: C.M. Collins Endowed Scholarship

## STUDENT MENTORING

- **Vanderbilt University**: Xiaohan Wang (Ph.D student)
- **University of Utah**: Ishrat Jahan Eliza (Ph.D student)
- **Kobe University**: Xing Huang (Ph.D student)
- **Arizona University**: Balla, Jake Dylan (Master student)