

# Guobin Shen

Room 208A

Phone: (+86) 139-3142-5808

Automation Mansion

Email: [shenguobin2021@ia.ac.cn](mailto:shenguobin2021@ia.ac.cn)

No. 95, Zhongguancun East Road, Haidian

Alt: [floyed\\_shen@outlook.com](mailto:floyed_shen@outlook.com)

Beijing, China

Homepage: [floyedshen.github.io](https://floyedshen.github.io)

## RESEARCH INTERESTS

---

I am passionate about integrating insights from neuroscience and cognitive science with advanced AI systems to build scalable, trustworthy, and safe AI. My research focuses on developing brain-inspired approaches to understand and improve large models, with emphasis on alignment methods, uncertainty quantification, and robustness against failure modes such as jailbreak attacks and hallucinations. I aim to build AI systems that are not only powerful but also scalable, interpretable, and reliable.

## ACADEMIC APPOINTMENTS

---

**Human Intelligence (Hi) Lab, RedNote**

Beijing, China

Ace Top Intern Program

September 2021 – Present

*Responsibilities:* Developed scalable AI alignment methods using efficient human feedback and automated preference learning.

## EDUCATION

---

**Institute of Automation, Chinese Academy of Sciences**

Beijing, China

Ph. D. in Machine Learning

September 2021 – June 2026 (expected)

*Advisor:* [Prof. Yi Zeng](#).

**Sun Yat-sen University**

Guangzhou, China

B. Eng. in Communication Engineering

September 2017 – June 2021

*Advisor:* [Prof. Xiang Chen](#).















*Grade:* 1/85 of graduating class.

## PUBLICATIONS





---

LLM Alignment & AI Safety:

1. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, He, Xiang, and Zeng, Yi. “Jailbreak Antidote: Runtime Safety-Utility Balance via Sparse Representation Adjustment in Large Language Models.” *Proceedings of the 13th International Conference on Learning Representations (ICLR)*, 2025. [\[OpenReview\]](#) [\[PDF\]](#)
2. **Shen, Guobin**, Zhao, Dongcheng, He, Xiang, Feng, Linghao, Dong, Yiting, Wang, Jihang, Zhang, Qian, and Zeng, Yi. “Neuro-Vision to Language: Image Reconstruction and





- Interaction via Non-invasive Brain Recordings.” *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS)*, 2024.  [\[PDF\]](#)  [\[Poster\]](#)
3. **Shen, Guobin**, Zhao, Dongcheng, Bao, Aorigele, He, Xiang, Dong, Yiting, and Zeng, Yi. “StressPrompt: Does Stress Impact Large Language Models and Human Performance Similarly?” *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI)*, 2025.  [\[OpenReview\]](#)  [\[PDF\]](#)
  4. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, Zhang, Qian, and Zeng, Yi. “Convergent Evolution across Modalities, Scales and Training Trajectories: Evidence for Human Brain-AI Alignment”, 2025.  [\[Arxiv\]](#)
  5. **Shen, Guobin**, Zhao, Dongcheng, Tong, Haibo, Li, Jindong, Zhao, Feifei, and Zeng, Yi. “Safety Instincts: LLMs Learn to Trust Their Internal Compass for Self-Defense.” *arXiv preprint arXiv:2510.01088*, 2025.  [\[Arxiv\]](#)
  6. **Shen, Guobin**, Zhao, Dongcheng, Feng, Linghao, He, Xiang, Wang, Jihang, Shen, Sicheng, Tong, Haibo, Dong, Yiting, Li, Jindong, Zheng, Xiang, and others. “PandaGuard: Systematic Evaluation of LLM Safety in the Era of Jailbreaking Attacks.” *arXiv preprint arXiv:2505.13862*, 2025.  [\[Project\]](#)  [\[Arxiv\]](#)  [\[Code\]](#)  [\[Dataset\]](#)
  7. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, Li, Yang, Li, Jindong, Sun, Kang, and Zeng, Yi. “Astrocyte-Enabled Advancements in Spiking Neural Networks for Large Language Modeling.” *arXiv preprint arXiv:2312.07625*, 2023.  [\[Arxiv\]](#)
  8. Wu, Ping, **Shen, Guobin**, Zhao, Dongcheng, Wang, Yuwei, Dong, Yiting, Shi, Yu, Lu, Enmeng, Zhao, Feifei, and Zeng, Yi. “CVC: A Large-Scale Chinese Value Rule Corpus for Value Alignment of Large Language Models.” *arXiv preprint arXiv:2506.01495*, 2025.  [\[Arxiv\]](#)  [\[Code\]](#)  [\[Dataset\]](#)

#### Spiking Neural Networks & Brain-Inspired AI:

9. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, and Zeng, Yi. “Brain-Inspired Neural Circuit Evolution for Spiking Neural Networks.” *Proceedings of the National Academy of Sciences*, vol. 120, no. 39, 2023, p. e2218173120. National Academy of Sciences.  [\[PDF\]](#)
10. **Shen, Guobin**, Zhao, Dongcheng, Li, Tenglong, Li, Jindong, and Zeng, Yi. “Are Conventional SNNs Really Efficient? A Perspective from Network Quantization.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 27538-27547.  [\[PDF\]](#)  [\[Poster\]](#)
11. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, Li, Yang, Zhao, Feifei, and Zeng, Yi. “Learning the Plasticity: Plasticity-Driven Learning Framework in Spiking Neural Networks.” *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.  [\[PDF\]](#)
12. **Shen, Guobin**, Zhao, Dongcheng, and Zeng, Yi. “Backpropagation with Biologically Plausible Spatiotemporal Adjustment for Training Deep Spiking Neural Networks.” *Patterns*, vol. 3, no. 6, 2022. Elsevier.  [\[PDF\]](#)
13. **Shen, Guobin**, Zhao, Dongcheng, and Zeng, Yi. “Exploiting Nonlinear Dendritic Adaptive Computation in Training Deep Spiking Neural Networks.” *Neural Networks*, vol. 170, 2024, pp. 190-201. Pergamon.  [\[PDF\]](#)
14. **Shen, Guobin**, Zhao, Dongcheng, and Zeng, Yi. “Exploiting High-Performance Spiking Neural Networks with Efficient Spiking Patterns.” *IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI)*, 2025.
15. **Shen, Guobin**, Zhao, Dongcheng, Shen, Sicheng, and Zeng, Yi. “Enhancing Spiking Transformers with Binary Attention Mechanisms.” *The Second Tiny Papers Track at ICLR 2024*.  [\[PDF\]](#)
16. **Shen, Guobin**, Zhao, Dongcheng, Dong, Yiting, Li, Yang, and Zeng, Yi. “Dive into the Power of Neuronal Heterogeneity.” *arXiv preprint arXiv:2305.11484*, 2023.  [\[Arxiv\]](#)




17. Zhao, Dongcheng, **Shen, Guobin**, Dong, Yiting, Li, Yang, and Zeng, Yi. “Improving Stability and Performance of Spiking Neural Networks through Enhancing Temporal Consistency.” *Pattern Recognition*, vol. 159, 2025, p. 111094. Pergamon.  [\[Arxiv\]](#)  [\[PDF\]](#)
18. Han, Bing, Zhao, Feifei, Zeng, Yi, and **Guobin Shen**. “Developmental Plasticity-Inspired Adaptive Pruning for Deep Spiking and Artificial Neural Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. IEEE.  [\[PDF\]](#)
19. Pan, Wenxuan, Zhao, Feifei, **Shen, Guobin**, Han, Bing, and Zeng, Yi. “Brain-Inspired Multi-Scale Evolutionary Neural Architecture Search for Deep Spiking Neural Networks.” *IEEE Transactions on Evolutionary Computation*, 2024. IEEE.
20. Shen, Sicheng, Zhao, Dongcheng, **Shen, Guobin**, and Zeng, Yi. “TIM: An Efficient Temporal Interaction Module for Spiking Transformer.” *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI 2024)*, 2024.  [\[PDF\]](#)
21. He, Xiang, Liu, Xiangxi, Li, Yang, Zhao, Dongcheng, **Shen, Guobin**, Kong, Qingqun, Yang, Xin, and Zeng, Yi. “CACE-Net: Co-guidance Attention and Contrastive Enhancement for Effective Audio-Visual Event Localization.” *Proceedings of the 32nd ACM International Conference on Multimedia (MM)*, 2024, pp. 985-993.  [\[OpenReview\]](#)  [\[PDF\]](#)
22. He, Xiang, Zhao, Dongcheng, Li, Yang, **Shen, Guobin**, Kong, Qingqun, and Zeng, Yi. “An Efficient Knowledge Transfer Strategy for Spiking Neural Networks from Static to Event Domain.” *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 38, no. 1, 2024, pp. 512-520.  [\[Arxiv\]](#)
23. Han, Bing, Zhao, Feifei, Zeng, Yi, Pan, Wenxuan, and **Shen, Guobin**. “Enhancing Efficient Continual Learning with Dynamic Structure Development of Spiking Neural Networks.” *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2023.  [\[PDF\]](#)
24. Zeng, Yi, Zhao, Dongcheng, Zhao, Feifei, **Shen, Guobin**, Dong, Yiting, Lu, Enmeng, Zhang, Qian, Sun, Yinqian, Liang, Qian, Zhao, Yuxuan, and others. “BrainCog: A Spiking Neural Network Based, Brain-Inspired Cognitive Intelligence Engine for Brain-Inspired AI and Brain Simulation.” *Patterns*, 2023, p. 100789.  [\[PDF\]](#)
25. Shen, Sicheng, Zhao, Dongcheng, Feng, Linghao, Yue, Zeyang, Li, Jindong, Li, Tenglong, **Shen, Guobin**, and Zeng, Yi. “STEP: A Unified Spiking Transformer Evaluation Platform for Fair and Reproducible Benchmarking.” *Advances in Neural Information Processing Systems (NeurIPS) Dataset and Benchmark Track*, 2025.  [\[PDF\]](#)

#### Hardware Acceleration & System Optimization:

26. **Shen, Guobin**, Li, Jindong, Li, Tenglong, Zhao, Dongcheng, and Zeng, Yi. “SpikePack: Enhanced Information Flow in Spiking Neural Networks with High Hardware Compatibility.” *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.  [\[PDF\]](#)
27. Li, Jindong, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. “Firefly: A High-Throughput Hardware Accelerator for Spiking Neural Networks with Efficient DSP and Memory Optimization.” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 31, no. 8, 2023, pp. 1178-1191. IEEE.  [\[PDF\]](#)
28. Li, Jindong, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. “Firefly v2: Advancing Hardware Support for High-Performance Spiking Neural Network with a Spatiotemporal FPGA Accelerator.” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2024. IEEE.  [\[PDF\]](#)
29. Li, Tenglong, Li, Jindong, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. “FireFly-S: Exploiting Dual-Side Sparsity for Spiking Neural Networks Acceleration with Reconfigurable Spatial Architecture.” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2024. IEEE.  [\[PDF\]](#)

30. Li, Jindong, Li, Tenglong, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. "Revealing Untapped DSP Optimization Potentials for FPGA-Based Systolic Matrix Engines." *2024 34th International Conference on Field-Programmable Logic and Applications (FPL)*, IEEE, 2024, pp. 197-203.  [\[Arxiv\]](#)  [\[PDF\]](#)
31. Li, Jindong, Li, Tenglong, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. "Pushing Up to the Limit of Memory Bandwidth and Capacity Utilization for Efficient LLM Decoding on Embedded FPGA." *2025 Design, Automation & Test in Europe Conference (DATE)*, IEEE, 2025, pp. 1-7.  [\[PDF\]](#)
32. Li, Jindong, Li, Tenglong, Chen, Ruiqi, **Shen, Guobin**, Zhao, Dongcheng, Zhang, Qian, and Zeng, Yi. "Hummingbird: A Smaller and Faster Large Language Model Accelerator on Embedded FPGA." *The 2025 International Conference on Computer-Aided Design (ICCAD)*, 2025.  [\[PDF\]](#)

#### Datasets & Data Augmentation:

33. **Shen, Guobin**, Zhao, Dongcheng, and Zeng, Yi. "EventMix: An Efficient Data Augmentation Strategy for Event-Based Learning." *Information Sciences*, vol. 644, 2023, p. 119170. Elsevier.  [\[PDF\]](#)
34. Dong, Yiting, He, Xiang, **Shen, Guobin**, Zhao, Dongcheng, Li, Yang, and Zeng, Yi. "EventZoom: A Progressive Approach to Event-Based Data Augmentation for Enhanced Neuromorphic Vision." *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI)*, 2025.  [\[OpenReview\]](#)
35. Dong, Yiting, Li, Yang, Zhao, Dongcheng, **Shen, Guobin**, and Zeng, Yi. "Bullying10K: A Large-Scale Neuromorphic Dataset Towards Privacy-Preserving Bullying Recognition." *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2024.  [\[PDF\]](#)

#### ACADEMIC SERVICES

---

Serve as a reviewer for conferences including *NeurIPS*, *ICML*, *ICLR*, *CVPR*, *ICCV*, *ECCV*, *AAAI*, *MM*, *AISTATS*, and journals including *IEEE Computational Intelligence Magazine*, *Pattern Recognition*, *Neural Networks*, and *Neurocomputing*.

#### TEACHING

---

University of Chinese Academy of Sciences

July 2023 – December 2023

Teaching Assistant, Systems and Computational Neuroscience

#### AWARDS AND HONORS

---

- **Best Paper Award for Chinese Scientists, *Cell Press* (2022)**
- **Best Paper Award, *Cell Press* (2023)**
- **Chinese Academy of Sciences President Scholarship (2025)**
  - Academic honor from Chinese Academy of Sciences, recognizing doctoral students with outstanding academic achievements (top 1%)
- **National Scholarship (Doctoral Student) (2024)**
  - Granted for exceptional research contributions and academic excellence (top 1%).
- **National Scholarship (Undergraduate) (2019, 2020)**
  - Awarded by the Chinese Government for outstanding performance in academics, extracurriculars, and leadership (top 2%).
- **National Second Prize, National Undergraduate Electronic Design Competition (2019)**
- **Runner-Up, International Aerial Robotics Competition (2019)**