Zihao Jing

zihaoj24@gmail.com | Homepage | Portfolio | Google Scholar

I am an AI researcher focused on LLM post-training, multimodal alignment, and model adaptation. I have three first-author papers at NeurIPS, ICLR, and ICML, along with hands-on industry experience in large-scale LLM training, embedding models, and production-oriented ML engineering at SenseTime.

SELECTED PUBLICATIONS

Top-Tier Conferences

[1] **Zihao Jing**, Qiuhao Zeng, Ruiyi Fang, Yan Yi Li, Yan Sun, Boyu Wang, Pingzhao Hu. *Scaling-Aware Adapter for Structure-Grounded LLM Reasoning.* **ICML 2026** [Code] [Model] [Data] [Poster] [Slides]

- Designed an entropy-guided adapter that selects instruction-relevant structural regions for variable-size inputs, addressing uniform treatment of positions in structure-grounded LLM reasoning. Achieved top-1 performance on 17 out of 18 structured reasoning benchmarks.

[2] **Zihao Jing**, Qiuhao Zeng, Ruiyi Fang, Yan Sun, Boyu Wang, Pingzhao Hu. *Entropy-Guided Dynamic Tokens for Graph–LLM Alignment in Molecular Understanding.* In **ICLR 2026**. [Code] [Model] [Data] [Poster] [Video][Slices]

- Built a dynamic graph-LLM connector that scales representation capacity based on input complexity, reducing structural information loss from fixed-size connectors. Evaluated in molecular understanding as a structured-input testbed, ranking #1 on 20/21 tasks and training 3.5× faster under equivalent settings.

[3] **Zihao Jing**, Yan Sun, Yan Yi Li, Sugitha Janarthanan, Alana Deng, Pingzhao Hu. *Structure-Aware Fusion with Progressive Injection for Multimodal Molecular Representation Learning.* In **NeurIPS 2025**. [Code] [Model] [Data]

- Developed a multimodal embedding model that stabilizes noisy structured inputs into reliable representations, improving robustness in multimodal embedding with geometric features. Ranked #1 on 22/29 interdisciplinary prediction tasks with up to 27% higher accuracy.

Selected Additional Works

[4] Junqin Huang, Zhongjie Hu, **Zihao Jing**, Mengya Gao, Yichao Wu. *Piccolo2: General Text Embedding with Multi-Task Hybrid Loss Training.* **SenseTime Technical Report, 2024** [Model] [Code]

- Developed training and evaluation pipelines for Piccolo2 (text embedding) in SenseNova team, supporting iterative optimization that achieved top-1 performance on C-MTEB benchmark in May 2024.

Additional Co-authored Publications

Co-authored 3 additional papers at **ICML 2026** and **ICLR 2026** (×2).

WORK EXPERIENCE

SenseTime *LLM Research Intern* 2023.09–2024.06

- Developed Piccolo-GPT, supporting text embedding and generation within a single LLM architecture.
- Built training & evaluation pipelines for text embedding models (Piccolo2), achieving top-1 on C-MTEB (May 2024).
- Fine-tuned a ∼100B LLM for vertical livestream marketing script generation at Sina Weibo.

Jina AI (acquired by Elastic) *AI Research Intern* 2023.04–2023.09

- Implemented LLM-based denoising and sentiment analysis pipeline for Budweiser; reduced API cost >13%.
- Led the evaluation and deployment of super-resolution models for commercialization (e.g., SwinIR).
- Integrated super-resolution inference pipelines into open-source LLM tooling in the LlamaIndex ecosystem.

EDUCATION

Western University, M.Sc. in Computer Science (Research-Based), Ontario, Canada Sep. 2024–Jun. 2026
Beihang University, B.Eng. in Software Engineering, Beijing, China Sep. 2019–Jun. 2024

RESEARCH GRANTS

Digital Research Alliance of Canada, RRG Competition 2026 — Secured 5×A100-80GB GPU-years on Canada's national supercomputing cluster (∼**US$80K** commercial value).

TECHNICAL SKILLS

- 3 years of experience (2023-2026) in embedding models, **multimodal LLM post-training (SFT, RL)**. Proficient in PyTorch, Transformers, PEFT, DeepSpeed, huggingFace, experiment tracking (W&B). [Portfolio]
- 3 years of experience in **Large-Scale Distributed Training**: Slurm, DeepSpeed, DDP/FSDP, HPC, cloud computing, Docker, Apptainer. [Training-Script]
- 5 years of experience (2021-2026) in **Software Engineering**: Strong in data structure and algorithms; strong coding skills in Linux/Shell, Docker, Python, C/C++, Java, SQL, and full-stack development. Led 3 large websites deployed in Shanghai G60, Beihang, and Chaoyang Production Department. [Full-Stack Project].
- Open-source contribution with 9+ models, 8+ datasets with 2k+ downloads overall. [HuggingFace] Experience with multi-agent systems ([MAS Project]), vector databases and RAG systems.