CAN JIN

cj574@scarletmail.rutgers.edu | Github | LinkedIn | Google Scholar Homepage
Rutgers University-New Brunswick, NJ

EDUCATION

Rutgers University New Brunswick, NJ
Ph.D. Candidate in Computer Science *Aug. 2024 – Present*

- Fully Funded Teaching and Research Assistant. Advisor: Prof. Dimitris Metaxas.
- **Research Interests:** Pre-training/Post-training/Inference of Foundation Models.

University of Science and Technology of China (USTC) Hefei, China
Master of Science in Mathematics *Sep. 2018 – Jul. 2020*

- Advisor: Prof. Jianliang Zhai. Focus: Stochastic Partial Differential Equations (SPDEs).

University of Science and Technology of China (USTC) Hefei, China
Bachelor of Science in Mathematics *Sep. 2014 – Jul. 2018*

- Advisor: Prof. Yi Wang. Focus: Probability Theory, Random Processes, Dynamical Systems.

RESEARCH EXPERIENCE

Adobe Incorporated San Jose / Remote, CA
Research Intern (Foundation Model Pre-training), Intern Paper Accepted to ICML2026 *2025*

- Identify critical limitations in existing Top-p routing for **Mixture-of-Experts** (MoE), specifically regarding uncontrolled sparsity and sensitivity to hyperparameter selection.
- Propose **DTop-p MoE**, a dynamic routing scheme that learns the probability threshold via a PI controller, introducing dynamic routing normalization to ensure expert diversity.
- Validate the robustness of DTop-p across diverse **scaling regimes**—including expert granularity, expert capacity, model size, and dataset size—consistently outperforming Top-k and Top-p baselines on NLP and CV tasks.

Massachusetts Institute of Technology Remote
Research Intern (Efficient AI), Intern Paper Accepted to AAAI2025 *2023*

- Conduct a pilot study on post-**pruning** visual prompts, revealing its inefficacy in improving the performance of well-fine-tuned **sparse vision models**.
- Develop **VPNs**, a novel data-model co-design paradigm that simultaneously optimizes weight masks and visual prompts for efficient sparse vision models.
- Achieve state-of-the-art efficiency-performance trade-offs across diverse datasets, architectures, and pruning regimes (LTH, OMP, GraSP, SNIP, SynFlow, HYDRA, and BiP), validated by extensive empirical results.

SELECTED PUBLICATIONS

[1] **Can Jin***, Hongwu Peng*, Dimitris N. Metaxas, et al. *Sparsity-Controllable Dynamic Top-p MoE for Large Foundation Model Pre-training.* **ICML 2026**. **[PDF]**

- *Propose DTop-p **MoE**, a dynamic routing mechanism that utilizes a Proportional-Integral controller and dynamic routing normalization to precisely control expert activation sparsity while adapting to varying token difficulty. DTop-p outperforms Top-k and Top-p MoE across **Large Language Models and Diffusion Transformers**.*

[2] **Can Jin***, Rui Wu*, Tong Che*, et al. *Reasoning over Precedents Alongside Statutes: Case-Augmented Deliberative Alignment for LLM Safety.* **ACL 2026 Main Conference.** **[PDF]**

- *Propose CADA, a case-augmented **deliberative alignment** framework that leverages **reinforcement learning** on self-generated reasoning chains to transition from rigid rule enforcement to flexible case-based reasoning, significantly reducing over-refusal while enhancing robustness against jailbreak attacks.*

[3] **Can Jin**, Jiakang Li, Rui Wu, Eddy Zhang, Dimitris N. Metaxas. *Weak Critics Make Strong Learners: On-Policy Critique Distillation for Scalable Oversight.* **ICML 2026 AI4MATH Workshop.** **[PDF]**

- *Propose weak-critic strong oversight and on-policy critique distillation (OPCD), showing that weak models can guide stronger models with useful critiques while improving reasoning and alignment under scalable oversight.*

[4] **Can Jin**, Tong Che, Dimitris N. Metaxas, et al. *LoR-VP: Low-Rank Visual Prompting For Efficient Vision Model Adaptation.* **ICLR 2025**. **[PDF]** **[Code]**

- *Design LoR-VP, a **low-rank visual prompting** technique for **efficient** vision model adaptation that reduces trainable parameters while outperforming full fine-tuning and standard visual prompting methods on object detection and segmentation benchmarks.*

[5] **Can Jin***, Tong Che*, Marco Parvone, et al. *Learning from Teaching Regularization: Generalizable Correlations Should Be Easy to Imitate.* **NeurIPS 2024**. **[PDF]** **[Code]**

- *Develop Learning from Teaching (LoT), a novel **regularization** technique for deep neural networks that enhances model **generalization** by training a teacher model to prioritize features that are easier for a student model to imitate, thereby filtering out spurious correlations.*

[6] **Can Jin***, Tianjin Huang*, Tianlong Chen, et al. *Visual Prompting Upgrades Neural Network Sparsification: A Data-Model Perspective.* **AAAI 2025**. **[PDF]** **[Code]**

- *Propose VPNs, a novel data-model co-design framework that simultaneously optimizes visual prompts and network **sparsity**, significantly enhancing the performance and transferability of sparse vision models.*

[7] **Can Jin***, Hongwu Peng*, Dimitris N. Metaxas, et al. *APEER: Automatic Prompt Engineering Enhances Large Language Model Reranking.* **WWW 2025 - RelWeb (Best Paper Award)**. **[PDF]**

- *Propose APEER, a novel automatic prompt engineering algorithm that iteratively generates and refines prompts to enhance the performance and transferability of Large Language Models in information retrieval reranking tasks.*

[8] **Can Jin**, Hongwu Peng, Dimitris N. Metaxas, et al. *Two heads are better than one: Test-time scaling of multi-agent collaborative reasoning.* **NeurIPS 2025 - SEA**. **[PDF]** **[Code]**

- *Develop MAS-TTS, a framework that integrates a specialized **multi-agent** training pipeline with an adaptive CEO agent to orchestrate **collaborative reasoning**, effectively optimizing **test-time scaling** for complex tasks.*

[9] **Can Jin***, Hongwu Peng*, Dimitris N. Metaxas, et al. *RankFlow: A Multi-Role Collaborative Reranking Workflow Utilizing Large Language Models.* **WWW 2025 - RelWeb**. **[PDF]**

- *Design RankFlow, an LLM-driven reranking framework that utilizes **multi-role collaboration** to enhance retrieval accuracy, demonstrating superior performance over existing baselines in extensive empirical studies.*

[10] Yang Zhou*, **Can Jin***, Zihan Dong, Zhepeng Wang, Yanting Yang, Shiyu Zhao, Lei Li, Runxue Bao, Yaochen Xie, Dimitris N. Metaxas. *DARE: Difficulty-Adaptive Reinforcement Learning with Co-Evolved Difficulty Estimation.* arXiv preprint. **[PDF]** **[Code]**

- *Introduce **DARE**, a difficulty-adaptive **Reinforcement Learning** framework that co-evolves policy-aligned difficulty estimation with dynamic data selection and difficulty-specific optimization, improving training efficiency, final accuracy, and inference-token efficiency for LLM reasoning.*

[11] **Can Jin**, Tong Che, Dimitris N. Metaxas, et al. *Your reward function for RL is your best PRM for search: Unifying RL and search-based TTS.* arXiv preprint. **[PDF]**

- *Introduce AIRL-S that unifies **Reinforcement Learning** and search-based **Test-Time Scaling**, demonstrating that RL reward functions can serve as optimal **Process Reward Models** for guiding search in complex reasoning tasks.*

Industry Experience

Meituan Dianping Corporation — Shanghai, China
Machine Learning Engineer (Full-time) — *2020 – 2023*

- **Sales Forecasting:** Develop XGBoost and DNN models with quantile prediction to forecast SKU sales, increasing prediction accuracy by **4%** (92% → 96%) and optimizing inventory management.
- **Time Forecasting:** Construct Attention-based DNNs to estimate warehouse task duration, improving accuracy by **9%** (74% → 83%) and significantly reducing labor expenditures and heightening workforce efficiency.

Skills & Awards

Honors: Best Paper Award (WWW 2025 RelWeb), Outstanding Provincial Graduate (USTC, Top 3%, 2020), Outstanding University Graduate (USTC, Top 15%, 2020), National Encouragement Scholarship (USTC, Top 15%, 2015).

Languages & Tools: Python, PyTorch, C/C++, SQL, Git, Slurm, Docker, Cursor, VS Code, RStudio, MATLAB.

Academic Service:

Teaching Assistant (Rutgers University):

- CS344: Algorithms (Spring 2026)
- CS211: Computer Architecture (Fall 2025)
- CS534: Computer Vision (Spring 2025)
- CS210: Data Management for Data Science (Fall 2024)

Conference Reviewer: NeurIPS 25/26, ICLR 25/26, ICML 24/26, CVPR 25/26, ECCV 26, AAAI 26, etc.

Journal Reviewer: Alexandria Engineering Journal, Information Fusion, Pattern Recognition, Signal Processing.