

Xiangjun (Ethan) Fu

xjf@seas.upenn.edu | (858) 539-5848 | Philadelphia, PA | github.com/EthanFusion03

EDUCATION

University of Pennsylvania — M.S.E. in Computer & Information Science Aug 2025 – May 2027
GPA: 3.95 / 4.0

University of California, San Diego — B.S. Mathematics–Computer Science Sep 2021 – Jun 2025
GPA: 3.9 / 4.0 | Provost Honors

Relevant Coursework: Statistical NLP, Deep Learning, Deep Generative Modeling, Reinforcement Learning, Machine Learning, Recommender Systems & Web Mining, Probability & Statistics.

RESEARCH EXPERIENCE

GMLR Lab, University of Pennsylvania — Graduate Researcher Aug 2025 – Present
Advised by Prof. Jiatao Gu

- Designing a continuous latent chain-of-thought reasoning pipeline on a Qwen3-8B backbone that operates in lower-dimensional space while preserving semantic structure.
- Fine-tuning an LLM-based VAE encoder to provide semantic guidance for latent thinking-token generation; using normalizing flows to autoregressively generate thinking tokens under an MSE objective against VAE-encoded latents.
- Computing step-wise log-likelihoods to enable reinforcement learning over reasoning trajectories, with evaluation protocols for trajectory selection.

McAuley Lab, UC San Diego — Undergraduate Researcher Dec 2024 – Dec 2025
Advised by Prof. Julian McAuley & PhD candidate Yupeng Hou

- Built an evaluation framework for benchmarking LLM text embeddings in retrieval and recommendation, including a public [Amazon dataset](#) with 2M+ training and 250K+ test instances for systematic comparison.
- Benchmarked embeddings from SimCSE, E5-Mistral-7B, text-embedding-3-large, Qwen-7B, Gemini, and OpenAI across collaborative filtering, sequential recommendation, and product search; co-first-authored the resulting ACL 2026 paper.

Shang Data Lab, UC San Diego — Undergraduate Researcher Sep 2023 – Jun 2024
Advised by Prof. Jingbo Shang & PhD candidate Letian Peng

- Refined weakly supervised text classification, reaching 93% peak accuracy on AG News (vs. 78% baseline) via KL-divergence matching between token-level text and class-prompt distributions.
- Implemented the framework in PyTorch with BERT, RoBERTa, and Gemma across AG News and Rotten Tomatoes.

PUBLICATIONS

- Yupeng Hou*, Jiacheng Li*, **Xiangjun Fu***, Zhankui He, An Yan, Xiusi Chen, Julian McAuley. "Bridging Language and Items for Retrieval and Recommendation: Benchmarking LLMs as Semantic Encoders." ACL 2026 Main Conference **Oral** (* equal contribution).
- Guancheng Tu*, **Xiangjun Fu***, Suhao Yu, Yao Tang, Haoqiang Kang, Lianhui Qin, Yizhe Zhang, Jiatao Gu. "Latent Reasoning with Normalizing Flows." Under Review. arXiv:2606.06447, 2026 (* equal contribution by coin toss).
- Xinhua Wang, **Xiangjun Fu**, Caibo Feng, Chunxiao Liu. "Larger Hausdorff Dimension in Scanning Pattern Facilitates Mamba-Based Methods in Low-Light Image Enhancement." Under Review. arXiv:2510.26001, 2025

TEACHING & COMMUNICATION

Academic Achievement Hub, UC San Diego — Supplemental Instruction Leader Jul 2024 – Jun 2025

- Led weekly workshops for 50+ students in Physics and Calculus (20 hrs/week), translating complex technical material into accessible, interactive lessons.
- Iterated lesson design based on student and faculty feedback; provided one-on-one mentorship and partnered with course instructors to align support with the curriculum.

SKILLS

Programming: Python, C/C++, Java, SQL, Bash

ML & NLP: PyTorch, Hugging Face Transformers, LLM fine-tuning (LoRA, full fine-tuning), normalizing flows, flow matching/diffusion, VAEs, embedding models (BERT, RoBERTa, Mistral, Qwen), OpenAI / Gemini / Claude APIs

Tools: NumPy, Pandas, Matplotlib, Jupyter, Git, Linux, Vim, gdb, SLURM / HPC clusters, conda

Languages: Fluent in English and Mandarin; conversational Cantonese and Spanish.