

Emily Xiao

emilyx@alumni.cmu.edu | Website | Google Scholar | LinkedIn | Twitter

EDUCATION

M.S. Carnegie Mellon University

Language Technologies Institute, School of Computer Science

Advisors: [Graham Neubig](#) and [Chenyan Xiong](#)

Aug 2024 – Dec 2025

GPA: 3.9/4.3

B.A. University of California, Berkeley

Major: Computer Science

Aug 2017 – May 2021

Major GPA: 3.7/4.0

PUBLICATIONS

Prompt-MII: Meta-Learning Instruction Induction for LLMs

Emily Xiao, Yixiao Zeng, Ada Chen, Chin-Jou Li, Amanda Bertsch, Graham Neubig

ICLR 2026

Efficient Many-Shot In-Context Learning with Dynamic Block-Sparse Attention

Emily Xiao, Chin-Jou Li, Yilin Zhang, Graham Neubig, Amanda Bertsch

ACL 2025 Main

DATE-LM: Benchmarking Data Attribution Evaluation for Large Language Models

Cathy Jiao*, Yijun Pan*, **Emily Xiao***, Daisy Sheng, Niket Jain, Hanzhang Zhao, Ishita Dasgupta, Jiaqi W. Ma,

Chenyan Xiong

NeurIPS 2025

In-context learning with long-context models: An in-depth exploration

Amanda Bertsch, Maor Ivgi, **Emily Xiao**, Uri Alon, Jonathan Berant, Matthew R Gormley, Graham Neubig

NAACL 2025 Main [[SAC Award for Language Modeling](#)]

Automatically generating cause-and-effect questions from passages

Katherine Stasaski, Manav Rathod, Tony Tu, **Emily Xiao**, Marti A Hearst

EACL 2021, BEA Workshop

ACADEMIC RESEARCH EXPERIENCE

LLM Task Adaptation

Fall 2024 – Present

CMU, Advised by [Graham Neubig](#)

- Analyzed behavior of long context in-context learning ([NAACL 2025](#)).
- Proposed efficient ICL using dynamic sparse attention; implemented with FlexAttention and transformers library modifications, achieving 2× speedup while recovering 95% accuracy ([ACL 2025](#)).
- Used RL to train an LLM for automatic prompt engineering; ran multi-node VeRL training using 3000+ datasets, achieving SOTA performance with 13× shorter prompt. ([ICLR 2026](#))

LLM Training Data Curation

Fall 2024 – Spring 2025

CMU, Advised by [Chenyan Xiong](#) and [Jiaqi Ma](#)

- Explored pre-training and post-training data selection with data attribution.
- Conducted 100+ continual pretraining runs to extract gradient and optimizer states for probing-based analysis, as well as pretraining from scratch with different data mixtures.
- Implemented efficient per-batch data selection using custom PyTorch hooks and modifications to LitGPT.
- Designed efficient eval framework and comprehensive analysis of existing methods ([NeurIPS 2025 D&B](#))

LLM Systems

Spring 2025

CMU, Course Project Advised by [Lei Li](#)

- Designed Transformer model with dynamic depth, achieving 1.4× faster inference compared to standard Transformer Baseline. (A+ project grade)

Synthetic Data Generation

Fall 2020

UC Berkeley, Advised by [Marti Hearst](#)

- Built synthetic data generation pipeline; finetuned BERT models; designed autoevals ([BEA@EACL 2021](#))

INDUSTRY EXPERIENCE

Research Scientist Resident, Bridgewater Associates

Present

[AIA Labs](#)

- Large language model post-training for finance.

Machine Learning Engineer, TikTok

Fall 2021 – Fall 2023

[Query Auto-Completion](#)

- Proposed new training-data and modeling method that mitigates position bias and click-baiting; Launched globally with +2% prediction accuracy, and gave talk at TikTok Search.
- Designed predictive pre-caching, making short-prefix responses 2× faster.
- End-to-end optimization for multi-stage retrieval and ranking.
- Applied NLP techniques to recommendation setting, including transformer-based query rewrite.

[Related Search](#)

- Sole developer of modeling and recommendation pipeline; drove 3× growth in search volume.

Software Engineer Intern, Instagram

Summer 2020

[Instagram Reels – Feed Ranking \(founding team\)](#)

- Developed a personalized short-video retrieval strategy; Feature-engineered CTR ranking model.

Founding Engineer, SuiteSocial

Fall 2019 - Spring 2020

[Startup building a brand-influencer matching platform](#)

Won 2nd Place at UC LAUNCH Accelerator Demo Day

- Built supervised model for brand-influencer affinity ranking.

Founding Engineer, Prelude

Spring 2019 - Summer 2019

[Startup building an automated event planning platform](#)

- Built the MVP; Contributed to business strategy and product research.

TEACHING/MENTORING

Lab Assistant

Fall 2018

Data Structures & Programming Methodology (CS 61B), UC Berkeley

Student Research Mentoring

2025

Ada Chen

CMU Undergraduate

2025

Hanzhang Zhao

CMU Master

ADDITIONAL

Languages: English, Chinese, Japanese, Spanish

Coding Languages: Python, C++, SQL, Java

Tools: PyTorch, TensorFlow, NumPy, Huggingface, vLLM, SGLang, LitGPT, Hadoop, Spark, Kafka, CUDA C++

CS Coursework: Data Structures; Probability and Statistics; CS Theory; Computer Architecture; Information Devices and Systems; Computer Security; UI/UX Design; Database Systems

ML Coursework: Machine Learning; NLP; Optimization Models; Advanced NLP; Deep Learning Systems; LLM Systems; Trustworthy AI Theory and Practice; Inference Algorithms for Language Modeling