FURONG JIA

jiafurong2002@gmail.com | flora-jia-jfr.github.io/furongjia.github.io/

Durham, North Carolina, United States

Research Interests

Natural Language Processing, Data-centric Machine Learning, Healthcare, Time Series, Multimodal Machine Learning

Education

- **Duke University** *Aug 2024 - Present*
  Doctor of Philosophy in Computer Science Durham, NC
  ◦ Advisor: Monica Agrawal GPA: 4.0/4.0

- **University of Southern California** *Aug 2020 - May 2024*
  Bachelor of Science in Computer Science & Bachelor of Science in Applied and Computational Mathematics Los Angeles, CA
  ◦ W.V.T. Rusch Engineering Honors Program, Viterbi Grand Challenges Scholar GPA: 3.96/4.0

Publications

Published:

[1] Interpreting Dataset Shift in Clinical Notes
ML4H 2025 (Machine Learning for Health Symposium)
Shariar Vaez-Ghaemi*, **Furong Jia***, Monica Agrawal.

[2] Counting Clues: A Lightweight Probabilistic Baseline Can Match an LLM
ML4H 2025 Findings (Machine Learning for Health Symposium)
Furong Jia*, Yuan Pu*, Finn Guo, Monica Agrawal.

[3] What Patients Really Ask: Exploring the Effect of False Assumptions in Patient Information Seeking
ML4H 2025 Findings (Machine Learning for Health Symposium)
Raymond M Xiong, **Furong Jia**, Lionel Wong, Monica Agrawal.

[4] Diagnosing our datasets: How does my language model understand clinical text?
CHIL 2025 (7th Annual Conference on Health, Inference, and Learning)
Furong Jia, David Sontag, Monica Agrawal.

[5] GPT4MTS: Prompt-based Large Language Model for Multimodal Time-series Forecasting.
EAAI-24 (The 14th Symposium on Educational Advances in Artificial Intelligence)
Furong Jia, Kevin Wang, Yixiang Zheng, Defu Cao, Yan Liu. (2024).

[6] TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting.
ICLR 2024 (The Twelfth International Conference on Learning Representations)
Defu Cao, **Furong Jia**, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, Yan Liu.

[7] I2I: Initializing Adapters with Improvised Knowledge.
CoLLAs 2023 (Second Conference on Lifelong Learning Agents)
Tejas Srinivasan, **Furong Jia**, Mohammad Rostami, Jesse Thomason.

Under Review:

[1] Batch-of-Thought: Cross-Instance Learning for Enhanced LLM Reasoning
Xuan Yang, **Furong Jia**, Roy Xie, Xi Xiong, Jian Li, Monica Agrawal

Research Experience

- **DukeNLP, Duke University** *Aug 2024 - Present*
  Advisor: Monica Agrawal Durham, NC
  ◦ Diagnosing our datasets: How does my language model learn clinical information?
    ∗ We examined the relationship of clinical information (clinical jargon understanding and unsupported medical claims) between the occurrence of keyword pairs in pretraining corpora and model performance. For clinical jargon understanding, we also curated a benchmark *Medlingo* using clinical abbreviations from real-world clinical notes. We further analyzed the sources of clinical information in pretraining datasets to guide future corpus design.
  ◦ Counting Clues: A Lightweight Probabilistic Baseline Can Match an LLM

    - We probed whether LLM success on clinical Multiple-Choice Diagnostic Questions reflects probabilistic inference by developing the *Frequency-Based Diagnostic Ranker (FBDR)*, a simple Naive Bayes approach that extracts clinical concepts and scores diagnosis options using co-occurrence statistics from pretraining corpora. We found that FBDR achieved accuracy closely matched by a corresponding 7B LLM using the same pretraining corpora, and the predictions are complementary. While LLM performance appears driven by mechanisms beyond simple frequency aggregation, a historically grounded, low-complexity, expert-system style approach still accounts for a substantial portion of benchmark performance.
  - Interpreting Dataset Shift in Clinical Notes
    - Distribution shift degrades ML performance, especially in clinical text. Therefore, actionability requires not just detection but *explanation*. We establish an extensible benchmark suite that induces synthetic distribution shifts using real clinical notes and develop two methods for assessing generated shift explanations. We further introduce **SIReNs**, a general-domain end-to-end approach that explains distributional differences by selecting representative notes from each. SIReNs reliably recover salient binary shifts with comparatively lower performance on subtle continuous changes, showing a gap to a ground-truth oracle and suggesting room for improvement in future methods.

- **Melady Lab, University of Southern California** *May 2023 - June 2024*
  Advisor: Yan Liu Los Angeles, CA
  - Tempo: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting
    - We developed TEMPO, a GPT architecture-based time series framework that combines trend/seasonal/residual decomposition with adaptive, selection-based prompting for non-stationary distribution shift. It achieves zero-shot state-of-the-art across diverse and unseen (including multi-modal) benchmarks, indicating strong potential as a foundation model for time series forecasting.
  - GPT4MTS: Prompt-based Large Language Model for Multimodal Time-series Forecasting
    - We introduced *GPT4MTS*, a prompt-based framework that fuses numeric time series with aligned textual context. We also built a GDELT-derived multimodal news impact dataset and showed consistent forecasting gains over strong unimodal baselines, demonstrating effective multimodal fusion and the value of extra-textual information.

- **GLAMOR Lab, University of Southern California** *Aug 2022 - Present*
  Advisor: Jesse Thomason Los Angeles, CA
  - I2I: Initializing Adapters with Improvised Knowledge
    - We applied parameter-efficient knowledge transfer using adapters in multimodal models for visual question-answering tasks. We implemented and trained model architectures on multiple vision-language tasks, including GQA, COCO-QA, and AQUA, to enhance forward transfer and reduce knowledge forgetting.

Work Experience

- **Galileo Financial Technologies, SoFi** *May 2022 - Aug. 2022*
  Software Engineering Intern US Remote
  - Independently designed, developed, and tested micro-services in Flask to support authentication and retrieve account information from the database, and generated scripts for Continuous Integration through Gitlab.
  - Independently designed, developed, and tested an account-information-component that creates web components to display an account information card using the lit-element framework in HTML/CSS and Typescript.

Honors and Awards

- **Albert Dorman Future Leader Award**, University of Southern California *2024*
- **Student Recognition Awards**, University of Southern California *2024*
- **Provost's Research Fellowship**, University of Southern California *Fall 2023*
- **CURVE Research Fellowship**, University of Southern California *Fall 2022, Spring 2023*
- **Academic Achievement Award Scholarship**, University of Southern California *2021*
- **ABC Innovation Prize**, University of Southern California *2021*

Service

- Teaching Assistant:
  - Applied Machine Learning, *Duke University* Spring 2025
  - Algorithms and Theory of Computing, *University of Southern California* Spring 2022
  - Web Publishing and Front-end Development, *University of Southern California* Fall 2021 - Spring 2024
- Reviewer:
  - ACL ARR 2024 (June, October), ACL ARR 2025 (February, May)
  - NeurIPS'25
  - ML4H'25