Kate Lin

kateslin99@gmail.com kateslin@github.io kateslin

Experience

Research Engineer *Sep 2021 - Current*
Google Research

Research Focus: Build Gemini evaluations that will guide Gemini to be more useful to humans

- Led the creation of an internal suite of social cognition evals for Gemini
- Drove the development of tabular reasoning evals for Gemini that are grounded in datasets published on the Web

Research Focus: Find, organize, and make use of datasets published on the Web

- Led project on answering natural language queries using data from tabular datasets published on the Web
- Drove synthetic query generation for improved dataset retrieval
- Led research project on analyzing, defining, and building methods to automatically infer relationships between datasets published on the Web. Resulting paper won a Best Paper Award.
- Designed, implemented, and launched a new feature to aggregate together all the versions of a dataset and present them together on the Dataset Search frontend
- Designed and implemented a solution for generating better training data for metadata extraction for tabular datasets, leading to 21% model precision improvement
- Improved dataset ranking results in Google Search by incorporating user interaction signals, leading to results triggering at a higher location for 78% of dataset related search queries

Student Researcher *May 2020 – Jan 2021*
Google Research

Research Focus: Develop automatic structured variational inference

- Developed and implemented new method of automatically structuring surrogate posteriors for variational inference that outperforms baseline surrogate posteriors methods (mean-field, multivariate normal, and normalizing flows). Paper accepted to AISTATS.

Undergraduate Research Assistant *Feb 2020 – Oct 2020*
MIT CSAIL Computer Aided Programming Group

Research Focus: Develop an interactive evaluation suite to understand how humans vs program synthesis models discover and replicate causal probabilistic programs
PI: Armando Solar-Lezama - **Direct Supervisor:** Zenna Tavares

- Designed and built Causal Inductive Synthesis Corpus (CISC), a suite of interactive problems designed for causal discovery for both agents & humans.
- Built v0 web interface with logging and replay functionalities for humans to interact with CISC eval
- Built Autumn to JavaScript transpiler to decrease latency of CISC interface.

Engineering Practicum Intern *May 2019 - Aug 2019*
Google Research

Research Focus: Devise a method to automatically generate crossword puzzles for specific topics trending in the news

- Fine tuned BERT to generate interesting crossword puzzle hints given Wikipedia articles and news articles
- Built pipeline to generate daily crossword puzzles based on trending news topics

Undergraduate Research Assistant *Sep 2018 - May 2019*
MIT Computational Cognitive Science Group

Research Focus: Develop inverse graphics model for visual processing of complex human poses in the brain
PI: Joshua B Tenenbaum - **Direct Supervisor:** Ilker Yildirim

- Modified SURREAL (Synthetic hUmans foR REAL tasks) generative model software to develop a suite of images of human bodies in different poses and locations that are difficult for both humans and computers to parse

Education

Wellesley College *2017-2021*
BA in Computer Science

Massachusetts Institute of Technology *2017-2021*
Cross-Registered Student

Oxford University - Worcester College *2019*
Visiting Student

Publications

Gemini 2.5 Pro Tech Report arXiv

Gemini Team

Tech Report

RADAR: Benchmarking Language Models on Imperfect Tabular Data arXiv

Ken Gu, Zhihan Zhang*, ***Kate Lin****, Yuwei Zhang*, Akshay Paruchuri*, Hong Yu*, Mehran Kazemi, Kumar Ayush, et al.

In Submission at NeurIPS

Relationships are Complicated! An Analysis of Relationships Between Datasets on the Web arXiv, Google Research Blog Post

Kate Lin, Tarfah Alrashed, Natasha Noy

International Semantic Web Conference (ISWC), 2024
Best Paper Award, Research Track

Automatic Structured Variational Inference arXiv

Luca Ambrogioni, ***Kate Lin***, Emily Fertig, Sharad Vikram, Max Hinne, Dave Moore, Marcel van Gerven

International Conference on Artificial Intelligence and Statistics (AISTATS), 2021

Causal Inductive Synthesis Corpus OpenReview

Zenna Tavares, Ria Das, Elizabeth Weeks, ***Kate Lin***, Joshua B Tenenbaum, Armando Solar-Lezama

NeurIPS Workshop on Computer Assisted Programming, 2020

Cross-Subject EEG Event-Related Potential Classification for Brain-Computer Interfaces Using Residual Networks HAL

Arnaldo Pereira, Dereck Padden, Jay Jantz, ***Kate Lin***, Ramses Alcaide-Aguirre

Tech Report

Invited Talks

Open Ecosystem For Dataset Discovery
Keynote Talk at DOME 4.0 Hackathon, 2024

Generating Topic Specific Crosswords from the News
Selected Intern Talk at Google Women Engineers Conference, 2019

Mentorship

Jerry Liu - Student Researcher 2025, PhD @ Columbia

Liri Fang - Student Researcher 2024, PhD @ UIUC

Samantha Dunn - STEP Intern 2023, now SWE @ Google Cloud

Nnamdi Obi - STEP Intern 2023, BS @ MIT