Avni Kothari

Research Engineer - AI Safety

- I am an AI Research Engineer with 5 years of software and ML engineering experience
- My AI research, presented at **ICLR, ICML, & NeurIPS**, focuses on: 1) improving transparency and alignment in high-stakes AI models & 2) ensuring fairness by detecting scenarios of preclusion
- I earned my M.S. in Computer Science as a DeepMind Fellow at UC San Diego

WORK EXPERIENCE

Research Engineer; University of California, SF; SF, CA — Sept 2023 – Present

- Developed a multimodal method to improve interpretability of black box models by generating and refining concept bottleneck models within a Bayesian framework for uncertainty quantification, outperforming baseline methods by 35%
- Created a method to use a Bayesian tree based model that merges Llama 2's domain knowledge with empirical data, increasing interpretability by 40%, as confirmed by domain experts, while matching top predictive performance and quantifying risk
- Designed and deployed a scalable ETL pipeline to process health record data, enabling ML training and evaluation for 3K+ patient records and 30K+ patient visits
- Built, deployed, and evaluated a custom clinical risk prediction model adopted by 10+ clinics, achieving 12% higher accuracy than the general clinical model
- Mentored 5+ peers and PhD students through teaching sessions and code reviews

Software Engineer; Edovo; Chicago, IL — June 2020 – May 2021

- Architected, tested, and deployed an educational content platform using Elasticsearch to handle 700K+ requests per day
- Led 10+ requirement gathering sessions with Product owners to re-build a platform
- Created a data pipeline and job to merge 4B rows of user event data in PostgreSQL

Lead Software Engineer; 8th Light; Chicago, IL — Jan 2017 – Mar 2019

- Implemented and deployed a scalable load testing platform simulating 1000+ RPS
- Engineered API integrations to sync 1000+ interactions/ minute in different timezones

PAPERS

Bayesian Concept Bottleneck Models with LLM Priors
Jean Feng, Avni Kothari, et al; *ICML under review, 2024*

Prediction Without Preclusion: Recourse Verification With Reachable Sets
Avni Kothari, et al; ***ICLR – Top 5% among submissions**, 2024*

Bayesian Priors From LLMs Make Clinical Prediction Models More Interpretable
Avni Kothari, et al; *AMIA – American Medical Informatics Association, Abstract, 2024*

Implementing a Predictive Model to Reduce Hospital Readmissions in a Safety Net Healthcare System
Arturo Gasga, Avni Kothari, et al; *ML4H - Machine Learning for Health, 2024* ***Oral Spotlight***

RESEARCH SOFTWARE

bc-llm — June 2024

- Implemented and developed a multimodal method using Metropolis Gibbs sampling to identify interpretable features for complex models
- Benchmarked and implemented 5+ comparator methods against our method, achieving performance comparable to or exceeding black-box models

reachml — June 2022

- Constructed a Mixed Integer Program to handle 50+ feature constraints for counterfactual explanations and test for robustness
- Created a model-agnostic fairness and safety audit to identify scenarios of preclusion
- Developed an HPC-based experimental pipeline to audit 200K+ individuals and benchmark results against baseline methods

CONTACT

- Oakland, CA
- +1-630-701-4490
- avni510@gmail.com

LinkedIn | Github
Website | Google Scholar

SKILLS

ML Engineering Skills:

- ML Pipelines and ETL
- Multimodal datasets
- ML Deployment
- ML Development & Evaluation

AI Safety Research Skills:

- Interpretability
- Robustness
- Fairness and Bias
- Risk Quantification
- LLMs and Foundation Models
- LLM Evaluations

Tools and Frameworks:

- Python (Hugging Face, Pytorch, Scikit-learn, Numpy, Pandas)
- AWS (EC2, S3, Terraform, Deployment Strategies)
- Elasticsearch
- DB: SQL, Postgres, DuckDB
- Docker

EDUCATION

University of California, San Diego
Masters in Computer Science
Thesis: Foundations of Model Agnostic Recourse Verification
San Diego, CA — 2021 - 2023

University of Texas at Austin
Bachelors in Mathematics and Economics
Minor: Computer Science
Austin, Texas — 2011 - 2016

POSTER PRESENTATIONS

- **NeurIPS** Workshop Statistical Foundations of LLMs and Foundation Models — Dec 2024
- **ICML** Workshop on Spurious Correlations, Invariance and Stability — Jul 2023
- **ICML** Workshop on Data-centric ML Research — Jul 2023

TEACHING EXPERIENCE

- Teaching Assistant for Interpretability & Explainability in ML — UC San Diego — Fall 2022