

# Avni Kothari

## AI / Machine Learning Engineer

I have 5+ years of experience building and deploying end-to-end production software and ML systems, taking products from initial conception through deployment to maintenance

### WORK EXPERIENCE

#### AI/ML Engineer; University of California, SF; CA Sept 2023 – Pres

- Architected and led end-to-end development of a GPT 4o-powered [patient summary](#) system in collaboration with 5+ clinicians and social workers, generating concise summaries for patients with 20+ (20000+ tokens) clinical notes to improve patient outcomes
- Established and led human annotation pipelines with social workers to validate LLM-generated summaries, then implemented automated evaluation frameworks with LLM-as-a-judge to scale across 100+ patient records
- Developed and authored a [publication](#) for a multimodal ML pipeline to integrate tabular and unstructured clinical data to improve black-box model interpretability, delivering 7% AUC improvement in readmission risk prediction
- Created and presented a hybrid LLM-empirical framework combining Llama 2's clinical knowledge with EHR data, achieving 40% interpretability improvement (validated by clinical experts) while matching top predictive performance and enabling production deployment
- 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
- Engineered, deployed, and evaluated a custom clinical risk prediction model adopted by 10+ clinics, achieving 12% higher accuracy than the general clinical model

#### ML Engineer; University of California, SD; CA June 2022 – Mar 2024

- Designed and developed a model agnostic auditing method and procedure to detect scenarios of bias for decision making models in high stakes settings
- Experimented with different models and datasets to identify scenarios of bias for up to 30% of a lending dataset
- Authored a [publication](#) that received an ICLR Spotlight (Top 5% among submissions)
- Collaborated with a lending expert to define 25+ real-world constraints on lending datasets to identify individuals subject to preclusion

#### 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 to thousands of users
- 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
- Developed a diabetes management iOS app to connect 60+ patients with diabetic nurses
- Mentored peers and residents through pair programming sessions and code reviews

### OPEN SOURCE SOFTWARE

#### [bc-llm](#) June 2024

- Implemented and developed a multimodal method using Bayesian methods 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 to test for robustness
- Created a model-agnostic fairness and safety audit to identify scenarios of preclusion

### CONTACT

- Oakland, CA
- +1-630-701-4490
- [avni510@gmail.com](mailto:avni510@gmail.com)
- [LinkedIn](#) | [Github](#)
- [Website](#) | [Google Scholar](#)

### SKILLS

#### ML Engineering Skills:

- ML Pipelines and ETL
- LLMs and Foundation Models
- LLM Evaluations
- Human Annotation
- Multimodal datasets
- ML Deployment
- ML Development & Evaluation
- AI Research

#### Tools and Frameworks:

- Python (Hugging Face, LangChain, Pytorch, Scikit-learn, Numpy, Pandas)
- AWS (EC2, S3, Terraform, Deployment Strategies)
- Elasticsearch
- DB: SQL, Postgres, DuckDB
- Docker
- Full Stack: Java, React, Typescript, Flask

### 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

### AI/ML RESEARCH 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



