

Rohit Mishra

rohitnmishra2gmail.com · (510)-449-2031 · github.com/ro-mish · linkedin.com/in/ro-mish · ro-mish.github.io

Education

Northwestern University	2024-2025
<i>M.S. in Data Science, concentration in AI</i>	
University of California, San Diego	2019-2023
<i>B.S. in Data Science, Minor in Entrepreneurship & Innovation</i>	

Professional Experience

Data Scientist	August 2023-Present
<i>Toyota Connected</i>	

- **Core developer of AI RAG Application Development for vehicle dealerships:** implemented advanced retrieval techniques and improved search relevancy accuracy by 45%, developed multi-modal data ingestion pipelines, and designed and owned application testing framework codebase, integrating ML Ops tools for thorough LLM observability
- Deployed LLM's and embeddings models on AWS, optimized and maintained endpoints, reduced inference latency by 96%, improved model throughput by 32x through multi-threading and utilizing GPU, and designed data governance strategy
- Collaborated across Toyota product teams to define LLM persona and consolidate fine-tuning data for prompt-engineering
- **Core contributor to Battery Degradation prediction models:** Researched and implemented various Time-Series Forecasting models for battery degradation, optimized hyperparameters, built ML pipeline with open source tools, reduced ETL and Inference latency by 88%, and saved 5 figures of development cost monthly
- Led and coordinated monthly, cross-team research paper literature reviews for emerging technologies discovery and adoption

Undergraduate Deep Learning Researcher	April 2023-December 2023
<i>Belkin Lab, Halıcıoğlu Data Science Institute at UC San Diego</i>	

- **Researched new feature extraction methods for Natural Language and Audio modeling** at Prof. Mikhail Belkin's lab
- Accelerated text embedding experiment code, achieving appx 70% speedup on high performance computing GPU clusters
- Collaborated with Postdoctoral Researchers across UCSD and Harvard, participating in literature reviews

Machine Learning Software Engineer Intern	June 2022-September 2022
<i>Toyota Connected</i>	

- **Implemented driver behavior modeling for Carbon Dioxide emissions reduction:** architected scalable ML pipeline for deployed model endpoints for extracting vehicle analytics and improved throughput by 2x
- Leveraged team's prior research of chemical reactions in vehicle emissions toward building Regression-based Time Series Models, achieving over 94% test prediction accuracy
- Optimized training processes for large datasets, improving development times by 83% and boosting overall compute efficiency
- **Drunk Driving Detection Computer Vision model:** Won 2nd place at Toyota hackathon, pitched business value to Executives

Software Engineer Intern	June 2021-September 2021
<i>Western Union</i>	

- **Contributed to Scheduled Payment Notification Service:** improved transaction times by 30% for 10 million users
- Boosted service reliability, wrote unit tests and achieved 95% line coverage, reducing post-release bugs by appx 40%
- Contributed to the modernization of legacy codebase to a microservice architecture and collaborated during design reviews

Patents & Publications

Driving Modification to Decrease Carbon	Pub. Date: Jul. 18, 2024
<i>US 20240242545 A1</i>	
Modeling Driver Style to Lower a Carbon Footprint	Pub. Date: Jun. 27, 2024
<i>US 20240211964 A1</i>	

Projects

• On Feature Scaling of Recursive Feature Machines	arxiv:2303.15745
• Pic2Prose: Open-Source NLP Package for Optical Character Recognition	Github/pic2prose
• PSA Grader: Data Science Scholarship Project	Poster

Certifications

- **AI Professional Program at Stanford University:** Deep Multi-Task and Meta Learning (Aug 2024), Machine Learning with Graphs (Jun 2024)
- **Google:** Production Machine Learning Systems (May 2024), Build and Deploy ML Solutions on Vertex AI (May 2024)

Skills

- **Machine Learning:** LLM Fine-tuning, RAG, Time-Series Forecasting, Predictive Modeling, ML Ops
- **Deep Learning:** PyTorch, TensorFlow, CUDA, Natural Language Processing, Audio Processing
- **Cloud & Big Data:** AWS (SageMaker, Glue, Bedrock, Lambda, Managed Airflow), Distributed Computing, Data Pipelines
- **Software Engineering:** Python, JavaScript, REST APIs, Microservices, Test-Driven Development
- **Data Science:** Feature Extraction, Embedding Techniques, Statistical Analysis
- **Tools & Frameworks:** Git, Docker, Kubernetes, Airflow, Spark, React.js