Pratik Kayal

Machine Learning Scientist II

pk2871@nyu.edu
https://www.linkedin.com/in/pratikkayal/
https://github.com/pratikkayal
+31643547600
1051KC, Amsterdam, NL

A Machine Learning Scientist who brings deep and cross-disciplinary skills in developing state-of-the-art machine learning solutions/deployment into production environments (+1M users/350+github stars), excellent research skill (NLP, LLM & Optimization expert/700+ citations/Primary author-ACL/NeurIPS), a background in Computer Science (IIT) and Artificial Intelligence (UvA), excellent programming skills (python), ability to solve complex problems/product requirements by applying approaches and methods from a variety of ML/AI disciplines, work cross-functionally and drive business value.

Work Experience

Machine Learning Engineer
Oct 2024 - Present

Rabobank | Amsterdam

Enterprise-scale fraud detection system (executes +1M data points weekly|large-scale machine learning model)

- Reduced false positive alerts by 32% through optimization of MoE-based Financial Crime classification model, saving accounting team 20 hours weekly, creating immediate and long-term business value
- Reduced model training costs by 230K annually by optimizing data processing pipeline (58% efficiency gain), feature exploration, and improving data quality with weekly rather than monthly model/data analysis updates, improving model trust and safety ratings
- Implemented MLOps coding practices (automated retraining, drift detection, active learning, CI/CD pipeline, full productionisation), increased deployment frequency by 400% and enabled the team to deliver 7 features

Machine Learning Scientist
Aug 2020 - Aug 2024

New York University

Enable LLMs to better understand video|First author LLM publication at A* NLP/AI/ML conference (ACL)

- Led the creation of sophisticated NLP-based frameworks, significantly advancing LLM interpretability and capability for video understanding tasks, directly contributing to state-of-the-art research at top-tier ML conferences.
- Fine-tune and engineered deep neural networks to work with LLMs, surpassing the best model by more than 17% on video understanding tasks, utilizing RAG for identifying and using the most appropriate data sources and LLM agents in critical steps

Research Engineer
Dec 2021 - Aug 2022

Rephrase AI | Bangalore

- Delivered a multi-domain scalable speech synthesis model (using Mel Spectrograms) handling 2K daily requests, and increasing audio quality by 67%, which enhanced client adoption rates by 18% (A/B testing/business value).
- Established technical roadmap and MLOps practices, achieving 99.5% model serving reliability

Research Engineer
Dec 2020 - Dec 2021

Deepen AI | Hyderabad

- Improved 3D bounding box detection accuracy by 43% (MaP) through engineering full lidar path and leveraging data
- Packaged models into Docker containers and implemented CI/CD for automated deployment
- Built multi-stage inference pipeline on Azure handling 500 simultaneous requests for 3D trajectory prediction that increased productivity of manual labeling methods by 84%

Core Skills

TensorFlow, Scikit-learn, AWS, Docker, AI/ML, Python, Redis, Machine Learning algorithms, Python, Data Structures, clear communication, data processing, SQL, Spark, Coding, Active Learning, Optimization, Big Data technologies, modeling techniques, feature exploration, evaluate model's performance, performance marketing, statistics, snowflake, mysql, experimental design, data manipulation, applied research, data governance, coach others, cassandra, dynamodb, pytorch, apache, pandas, numpy, Java, natural language processing, deep learning, research and development, Prototyping, ideation, physics, mathematics, benchmarking, data analytics, data generation, large data sets, data visualization, Software Development, continuous improvement

Education

University of Amsterdam

Aug 2022 - Feb 2024

Master of Science (MSc) Artificial Intelligence

Indian Institute of Technology (IIT) Gandhinagar

Aug 2016 - Aug 2020

Bachelors (B.Tech / BSc) Computer Science and Engineering

Publications

Large Language Models Are Natural Video Popularity Predictors

Association for Computational Linguistics (ACL) 2025

AutoCoder: Leveraging Transformers for Automatic Code Synthesis

Conference on Neural Information Processing Systems (NeurIPS/AIRPLANS)

Tables to LaTeX : Structure and Content Extraction from Scientific Tables

International Journal on Document Analysis and Recognition (IJDAR)