Zsolt Pajor-Gyulai

Data Scientist, Machine Learning Engineer

Permanent Resident of the United States

pgyzs1@gmail.com

240-330-3718

Montclair, NJ (NYC Metro Area)

Website; Github; Arxiv, Google Scholar

Cruise LLC - Autonomous Vehicles

Staff Machine Learning Engineer, AI Search **July 2024 – Present**

Led a team of 6 people developing natural language-based search tools that combine vector search based on embeddings from internal/external multimodal foundational models and heuristic mining. This tool discovered long tail, low coverage AV data supporting training and evaluation of AV models.

- Technical direction setting through cross-functional collaboration to optimally serve the needs of our Perception, Behavior planning, and Safety Evaluation partners by creating searchable embeddings of various modalities.
- Designed and implemented an incremental, daily data processing and inference pipeline unlocking continuous ingestion into our search corpus, supporting continuous update of AV model datasets.

Staff Data Scientist, AV Performance Evaluation Data Science **September 2023 – July 2024**

Technical leadership for the organization. Strategic direction setting, mentorship, and cross-functional alignment for an organization of 20 scientists/engineers evaluating AV performance on road and in simulation.

- Developed ML-based Safety Metrics strategy and drove strategic alignment of the related metric ecosystem. Initiated and prototyped a monitoring system based on first principle leading indicators of risky behavior.
- Received equivalent of *Exceeds Expectations* rating for clarifying anddriving the roadmap of our safety metrics ecosystem towards measuring AV Performance in alignment with the larger business strategy.

Senior Data Scientist, Data Science **January 2022 - September 2023**

Developed an end-to-end ML system running live on the autonomy stack for collision risk detection without reliance on the signal from a safety driver.

- Prototyped, designed, trained and validated the initial product and launched it on the autonomy stack from scratch (C++).
- Led a team of 4 scientists/engineers to improve the model's lift by a factor of ~8-10x and to build out several analytics capabilities providing insights on AV safety to the customer.
- Received *Exceeds Expectation* rating.

The Voleon Group - Quantitative Finance

Machine Learning Researcher (Promoted in 2019) **September 2018 – January 2022**

Developed scalable, production-grade ML systems for systematic trading, leveraging cutting-edge AI, deep neural networks, and tree models for stock return prediction.

- Built and maintained multi billion AUM market-neutral stratarb strategies across Asia-Pacific, European equities, and US futures with state-of-the-art planning, optimization, and risk management.
- Engineered optimized data pipelines and deployed forecasters and portfolio planners to production, resolving operational challenges.
- Managed the entire ML lifecycle, including data pipelines, feature engineering, model training, validation, and productization for live inference.

CIMS, New York University - Academic Research

Assistant Professor/Courant Instructor **September 2015 – June 2018**

Studied the long-time behavior of diffusions along heteroclinic networks - dynamical systems with collections of hyperbolic saddle points and connecting heteroclinic orbits.

- Published 5 scientific articles (13 total) in high-prestige journals (see arxiv page) including a 140+ page magnum opus establishing the presence of polynomial time-scale metastability for such systems.

Education

2015 June	Ph.D. in Mathematics (Probability)	University of Maryland College Park
2010 June	B.Sc. and M.Sc. in Physics	Budapest University of Technology and Economics

Technical expertise

Programming Languages and libraries

(Python) Numpy, Pandas, Scikit-Learn, XGBoost, PyTorch; (R) data.table, tidyverse; (C++) ROS

Public Infrastructure Frameworks

(Cloud) Google Cloud, BigQuery, DBT; (Tools) Linux, Docker