Rishav Aryan

Pleasanton CA | +1 571 604 8902 | rishavaryan058@gmail.com | LinkedIn | Github | Website & Papers

Summary

Applied ML Engineer with experience building autonomous LLM-based agents for real-world systems. Worked on multi-agent workflows, reinforcement learning for preference optimization, retrieval-augmented reasoning, and simulation-based evaluation. First author on an ACL 2026 submission on robust cross-platform agents. Strong Python engineer with product intuition, focused on deploying agentic systems from research to production.

Technologies

- **Programming Languages**: Python
- **Machine Learning & Modeling**: Transformers, Attention Mechanisms, Reinforcement Learning (PPO), VAEs, CNNs, GMMs, Representation Learning
- **Multimodal & LLM Systems**: Vision-Language Models, Embodied Agents, Retrieval-Augmented Generation (RAG), Semantic Search, AI Systems
- **Mathematics & Theory**: Linear Algebra, Probability, Statistics, Optimization
- **Systems & Implementation**: PyTorch, TensorFlow, FAISS, FastAPI, Docker, AWS (Lambda, EC2, S3), Software Engineering
- **Research Practice**: Problem Framing, End-to-End ML Pipelines, Ablation Studies, Robustness Evaluation, Curiosity

Education

George Mason University **Aug 2023 - May 2025**

Masters, Data Analytics and Engineering

- **Coursework:** Applied Deep Learning, Big Data to Information, Applied Predictive Analytics, Derivative Products and Risk Management

Vellore Institute of Technology (India) **Jun 2019 - Jul 2023**

Bachelors, Electronics and Communication Engineering

- **Coursework:** Neural Network and Deep Learning, Probability Theory and Random Process, Advanced Engineering Calculas, Data Structures and Algorithm

Experience

George Mason University **Jul 2025 - Present**

Research Assistant --- Machine Learning & Reinforcement Learning *United States (remote)*

- Designed a regime-aware reinforcement learning system to address non-stationarity in time-series data.
- Learned latent state representations using VAE + GMM to provide structured context for policy learning.
- Implemented hierarchical PPO-based allocation policies with realistic constraints and bias-free evaluation.
- Achieved consistent improvements in risk-adjusted performance over static baselines, reaching Sharpe ≈ 1.23, Sortino ≈ 2.01, Calmar ≈ 1.60, and positive alpha (up to +0.23) across test windows.

Wall Street Quants **Jul 2024 - Oct 2024**

Quantitative Research Intern *New York (Remote)*

- Developed strategies for crypto markets that balanced high returns and volatility risk.
- Enhanced profitability of trading models by optimizing strategies for the top 10 cryptocurrencies using python.
- Designed momentum and reversal strategies, calibrated RSI/moving averages, and applied volatility filters across simulated scenarios.
- Boosted annualized returns from 18% to 25%, raised Sortino ratio to 1.85, and improved Sharpe ratio by 15%.

Foxmula **May 2022 - Jul 2022**

Machine Learning Intern *India*

- Collaborated with the HR analytics team to develop models using TensorFlow and AWS, predicting promotions and identifying dissatisfaction drivers, enhancing HR decision-making.
- Built accurate, interpretable ML models with custom features, enhancing model reliability and supporting strategic HR initiatives.
- Implemented ensemble models (XGBoost, Stacking), engineered tenure + skill-gap features, and automated ML pipelines.
- Achieved 90% precision, improved accuracy by 15%, and saved 10+ hours weekly while informing HR strategy.

Projects

iOS World Agents - Embodied AI Evaluation Framework | Link

- Built an embodied simulation framework to evaluate LLM-driven agents performing real iOS actions in simulator environments, enabling behavioral evaluation beyond static benchmarks.
- Designed JSON-based task schemas and evaluation pipelines covering 50+ multi-step tasks across Safari, Maps, Calendar, Files, and Settings.
- Orchestrated GPT-4o, Gemini-1.5, and Grok-2 agents to enable controlled cross-model behavioral comparison under identical simulated conditions.
- Developed reflexion and TextGrad feedback loops, improving task completion and recovery behavior by ~8–10% without fine-tuning.

NetGuard - Agentic AI for Cybersecurity | Link

- Led the design and implementation of NetGuard, a serverless multi-agent cybersecurity system powered by SecureGPT and AWS Lambda, achieving 94.5% accuracy and 90.8% F1-score in real-time network threat detection and compliance enforcement.
- Developed modular AI agents (Ingestor, Analyzer, Aggregator, Ticket Generator) to automate network traffic analysis, SecureGPT-based threat classification, and NIST-compliant ticket generation integrated with Jira and Cyber Vision dashboards.
- Engineered synthetic datasets and prompt-driven LLM pipelines for anomaly log enrichment, enabling explainable AI decisions and reflexion-based system self-improvement, validated across 6,000+ data points with minimal latency.

ConfluBot - AI Assistant for Confluence | Link

- Designed an AI-powered semantic search assistant to improve the search of documentation for Confluence users.
- Implemented sentence transformers (all-MiniLM-L6-v2) with FAISS for vector retrieval, wrapped in FastAPI for real-time access.
- Deployed AI-powered Confluence assistant, boosting engineering team efficiency after achieving 85%+ retrieval precision.

Publications

- First Author, ACL 2026 Submission.Learning to Act Anywhere: Experience-Based Similarity for Universal Interface Agents..*training-free cross-platform UI agents using FAISS-based Elastic Visual Memory, achieving ~40% higher task success under interface perturbations with 83 ms per-step latency.*
- Co-Author, NLDB 2025.Multimodal Event Detection: Current Approaches and Defining the New Playground through LLMs and VLMs..*systematic evaluation of multimodal ED identifying failure modes of generative LLM/VLM approaches and conditions where supervised fusion models outperform.*