VIRAT SRIVASTAVA

+91-9893152542 | Pune, India | viratsrivastava.work@gmail.com
linkedin.com/in/virat-srivastava | github.com/ViratSrivastava | viratsrivastava-portfolio.vercel.app

Work Experience

Agentic AI Engineer **Aug 2025 - Present**
FirstEigen LLC *Pune, India*
Project: BuckGPT Development

- Designed **BuckGPT v2 intent classification architecture** with **12 modular, intent-specific pipelines**, supporting dynamic preprocessing, embedding selection, entity extraction, and rule-book–driven actions
- Built **scalable conversational state infrastructure** using **Redis–PostgreSQL hot–cold architecture**, enabling low-latency active context, durable archival, analytics, and full auditability
- **Reduced AI summary generation and retrieval latency from ∼20s to ∼20ms** by designing database-backed caching, timestamp-aware regeneration logic, and optimized execution paths

Software Development Engineer Intern **Sep 2024 - Oct 2024**
National Remote Sensing Center, ISRO *Hyderabad, India*
Project: Renewable Solar Energy Estimation in Urban Areas Using Satellite Images

- Developed 3 deep object detection and classification models for urban development analysis across spatial images ranging from 0.30m to 20m resolution
- Processed 3.5 million square kilometres of satellite data with 99%+ pipeline reliability by building robust Data Processing Frameworks handling multi-resolution imagery
- Improved satellite image classification accuracy to 97-99% from baseline models by engineering comprehensive data pipelines and implementing custom deep learning architectures for urban development analysis

Deep Learning Engineer & Founder **Oct 2023 - Aug 2024**
AntiProton Labs *Bhopal, India (Hybrid)*
Project: AI-Driven Hexapod Robot System

- Developed modular software pipeline reducing development time by 40% by integrating ROS with reinforcement learning controllers for autonomous gait planning and energy optimization in 18-DOF hexapod robot
- Achieved 95%+ locomotion success rate across 5 diverse terrain types by designing and training deep RL models (PPO, TRPO) for autonomous obstacle avoidance and terrain adaptation
- Reduced model validation time by 60% across 100+ control policy iterations by building comprehensive simulation-testing pipeline using ISAAC Sim and custom physics layers

Technical Skills

Programming Languages: Python, C++, SQL, JavaScript, REST APIs,, FastAPI, HTML/CSS
Specialized Areas: NLP, Generative AI, LLM, AI, Data Analysis, Deep Learning, CUDA
ML/DL Frameworks: PyTorch, TensorFlow, Keras, Scikit-Learn, LangGraph, LangChain
Databases & Cloud: MySQL, PostgreSQL, AWS, Docker, Kubernetes, Machine Learning Pipelines, Model Deployment

Key Projects

Trade Agents - Automated Trading System *GitHub*
Quantitative Analyst | Machine Learning Engineer

- **Developed Deep Q-Network (DQN) stock trading system with 59-dimensional state space and 54 technical indicators** processing **10+ stock symbols** (AAPL, AMZN, GOOGL, MSFT, TSLA) using **PyTorch, TA-Lib**, and custom reinforcement learning algorithms **trained models over 1,500+ episodes per stock symbol with GPU acceleration**
- **Implemented comprehensive feature engineering pipeline computing 54 technical indicators** including SMA, EMA, RSI, MACD, BB, ATR, and OBV with **MinMaxScaler normalization** for multi-source financial data integration

Remma-O1: Open-Source Language Model *GitHub*
ML Engineer | PyTorch Developer

- Architected 1.17B parameter transformer from scratch within 8GB VRAM constraints by implementing custom sparse-attention mechanisms and memory-efficient training pipeline for consumer hardware deployment
- **Built production-ready training infrastructure supporting distributed learning with automated checkpointing** by **developing comprehensive PyTorch framework** with gradient clipping, and cosine annealing scheduler
- **Engineered zero-dependency deployment system reducing setup complexity by 90%** by **creating modular architecture with YAML configuration management** and automated data preprocessing pipelines

Education & Certifications

Integrated Master of Technology in Artificial Intelligence (2021-2026)
Vellore Institute of Technology, Bhopal
Runner Up - Industrial Conclave VIT Bhopal 2024
Leadership: AI Club Secretary - organized 12 technical workshops / hackathons and built an intelligent precision farming system with 24% yield improvement through automated soil analysis and nutrient dispensing
Certifications: AWS Certified Cloud Practitioner