Varun Phanindra Shrivathsa

E-Mail: vphan@uic.edu | Mobile: (312)459-8769 | Chicago, IL |LinkedIn: linkedin.com/in/varun-p-shrivathsa/
GitHub:github.com/varunpshrivathsa |Medium: https://medium.com/@varunpshrivathsa

SUMMARY

MSCS Thesis student at UIC specializing in Machine Learning, Deep Learning, and Robotics. Experienced in building end-to-end ML pipelines, LLM fine-tuning, and recommendation systems using PyTorch, TensorFlow, Scikit-learn & Linux. Skilled in MLOps practices, cloud deployment (AWS, GCP, Kubernetes). Research includes robotic exploration with hands-on AgileX LIMO and SCOUT platforms.

SKILLS

Programming & Data: Python, C/C++, SQL, Pandas, NumPy, SciPy, JavaScript
Machine Learning: Scikit-learn, XGBoost, LightGBM, CatBoost, Statsmodels, H2O.ai, AutoML, Prophet
Deep Learning & GenAI: PyTorch, TensorFlow, Keras, ONNX, Hugging Face, Transformers, LLMs (GPT, LLaMA, Mistral), LangChain, LangGraph, OpenAI API, LoRA/QLoRA, JAX, Flax, DeepSpeed, Megatron-LM, PyTorch Lightning, OpenCV, spaCy
Frameworks & Systems: FastAPI, Flask, Node.js, React, GraphQL, gRPC, Kafka, Airflow, Spark, Redis
Data Engineering & Databases: PostgreSQL, MySQL, MongoDB, DynamoDB, Cassandra, Elasticsearch, Pinecone, FAISS
MLOps & Deployment: MLflow, Kubeflow, Ray Tune, Optuna, Docker, Docker Compose, Kubernetes, AWS (ECS, S3, Lambda, EKS), GCP (GKE, Vertex AI, Cloud Storage), Azure, Terraform, Helm, Git, GitHub Actions, Jenkins
Visualization, Monitoring & Testing: Prometheus, Grafana, Datadog, PyTest, Selenium, Matplotlib, Seaborn, Plotly, Power BI, Tableau
Autonomous Driving & Robotics: ROS, SLAM, LiDAR Sensor Fusion, HD Mapping, CARLA, NVIDIA Omniverse / DRIVE Sim
GPU Acceleration: CUDA, PyTorch CUDA, cuDNN, TensorRT, NCCL, OpenCL

EDUCATION

University of Illinois Chicago
Master of Science in Computer Science (Thesis: *Robotic Planning & Localization,* Advisor: Prof. Wenhao Luo) — Chicago, USA
Coursework: Machine Learning, Social Robotics, Data Science, Data Algorithmic Fairness, Advanced NLP — Present-Apr 2026

Dayananda Sagar University
Bachelor of Technology - Computer Science — Bengaluru, India
Coursework: Artificial Intelligence, Full Stack Development, Data Structures & Algorithms — Jun 2020 - Jun 2024

PROJECTS

RecomX: Scalable Multimodal Recommendation Framework
- Built a recommendation engine with collaborative filtering, content models, and multimodal embeddings (text, image, metadata).
- Trained multimodal encoders with PyTorch, Transformers, improving cold-start accuracy on sparse profiles and unseen items.
- Designed evaluation metrics (Hit Rate@K, NDCG), achieving 17% lift in ranking accuracy over baseline models on large datasets.
- Deployed FastAPI services and automated Airflow retraining with MLflow tracking, monitored via Prometheus/Grafana.

LLM-RAG Engine: Fine-Tuning & Retrieval-Augmented Generation Platform
- Fine-tuned LLaMA-2 & GPT3.5 (LoRA/QLoRA) and integrated RAG pipelines for improved factual grounding.
- Designed evaluation pipeline with BLEU, ROUGE & factuality scores, achieving 20% higher retrieval precision compared to baselines.
- Built GPU-accelerated embedding generation and FAISS vector search for efficient multi-document queries at scale.
- Deployed on AWS/GCP Kubernetes with FastAPI and GitHub Actions, with 5K+ concurrent queries & <150 ms latency and 99% uptime.

PipeLineX: End-to-End MLOps Orchestration Framework
- Created a distributed ML lifecycle for ingestion, preprocessing, training, evaluation, and deployment with reproducible workflows.
- Integrated MLflow tracking, Optuna/Ray Tune tuning, and Dockerized jobs via Kubernetes, Airflow, and Kubeflow pipelines.
- Built monitoring with Prometheus, Grafana, and ELK stack, enabling drift detection, anomaly alerts, and automated retraining triggers.
- Deployed on AWS S3/EKS and GCP Vertex AI with 99.8% reliability, 12% accuracy improvement via automated tuning.

RLPlan: Reinforcement Learning for Autonomous Driving in Simulation
- Built a planning module using RL in CARLA for lane changes, merging, and obstacle avoidance in dense traffic and adverse weather.
- Implemented PPO/DQN agents with reward functions balancing safety, comfort, surpassing rule-based and MPC planners.
- Trained with CUDA-accelerated PyTorch, achieving 3 times faster policy convergence and real-time rollout sampling.
- Validated in simulation, reaching 95% task success and reducing collisions by 18% compared to MPC baselines.

PUBLICATIONS

- **"LaneVectorNet: Lane Detection via Directional Vector Field Propagation."** arXiv preprint, Targeting submission to CVC 2026)
- **"Environmental Impact Analysis using Satellite Image Processing: A Case Study on Bangalore STRR Phase-1."** 2024 IEEE ASIANCON
- **"Deepfake Detection using LSTM and XResNet."** (IJRASET).[DOI: 10.22214/ijraset.2023.57056]

WORK EXPERIENCE

G19 Studio — Chicago, US (Remote)
ML Project Intern — May 2024 – July 2024
- Built 'TwinVerse', a human digital twin using real-time Apple Watch data with PyTorch, PPO-RL, and XGBoost for stress & fall detection.
- Deployed with multi-modal signal processing and personalized recovery recommendations as Dockerized FastAPI microservices with Redis caching, improving response latency by 40% under production-scale healthcare analytics workloads.

Mekhalyn — Bengaluru, India
Software Developer Intern — Jan 2024 – Apr 2024
- Led the development of a recruiter analytics platform with Flask, React, and PostgreSQL, deployed via Kubernetes and CI/CD.
- Fine-tuned GPT-3.5 using LoRA for resume insights, XGBoost with GPU inference to 5M+ records, reducing screening time by 38%.
- Implemented role-based authentication and RESTful APIs with Git-based workflows and unit testing, improving platform security.

National Institute of Advanced Studies (NIAS) — Bengaluru, India
Research Intern — **J**un 2023 – Jun 2024
- Led a team in developing a GIS pipeline for Impact analysis of Bengaluru STRR Phase-1 infrastructure project using data from multiple satellites - Cartosat, Landsat and ALOS-PALSAR to categorise 200m sqm with an accuracy of 92%.
- Trained 2D CNNs on multispectral bands for automated land-cover segmentation, improving accuracy in built-up detection.
- Obtained Karnataka State Council for Science and Technology(KSCST) Research Funding.