Tanmay Ranaware

San Jose, California | +1 408-549-6932 | tanmayranware14@gmail.com | LinkedIn Profile | Github Profile

Education

San Jose State University — San Jose, California
Masters of Science: Computer Engineering — *Aug. 2025 – Present*

National Institute Of Technology, Karnataka — India
Bachelors of Technology: Electrical And Electronics Engineering — *August 2019 – June 2023*

Technical Skills

Programming: Python, Java, C++, Bash
ML/AI: LLMs, Prompt Engineering, Generative AI, Transfer Learning, Synthetic Data, Alignment
Frameworks: TensorFlow, PyTorch, scikit-learn, Hugging Face Transformers
Data & Processing: Apache PySpark, Pandas, NumPy, Feature Engineering, Data Preprocessing
Deployment: FastAPI, Flask, REST/gRPC, A/B Testing, Model Deployment
MLOps Tools: Docker, Kubernetes, MLflow, Airflow, DVC, Weights & Biases, AWS, GCP, Terraform, Prometheus

Experience

1.Blackbuck(Zinka Logistics) - Software Engineer Full Time — July 2023 – June 2025
a.Financial Services Team — *Bangalore, India*

- Engineered **LLM-powered document analysis workflows**, reducing loan collection time by **40–50%** and showcasing applied ML for enterprise-scale decision-making.
- Architected account aggregator **RESTful APIs** (**Java, Spring Boot, OneMoney**) with **Circuit Breaker patterns** to sustain **99.9% uptime**, improving robustness in production ML pipelines.
- Applied rigorous **Test-Driven Development (90%+ coverage)**, strengthening model integration reliability and accelerating safe ML deployment cycles.

b.Gps Team

- Designed high-throughput **telemetry pipelines** (**Kafka, AWS Glue**) processing **680GB/day**, powering dashboards and ML-driven monitoring for **20000+ IoT devices**.
- Developed **backend microservices** (**Java, Spring Boot, MySQL, Redis**) for GPS lifecycles, cutting device checkout latency from **170s → 50s**, improving experimental turnaround.
- Resolved **200+ production incidents** by enhancing **observability** and fine-tuning **Kubernetes autoscaling (HPA/Cluster Autoscaler) and liveness/readiness probes**, achieving a sustained **99.99% SLA** while driving cross-functional collaboration under high-pressure conditions.

2.Legato Health Technologies — May 2022 – July 2022
Software Engineer Intern — *Bangalore, India*

- Built secure **synthetic data pipelines** (**Python, Hive, CTGAN**), enabling privacy-preserving ML experimentation and reducing reliance on sensitive production datasets.
- Automated distributed telemetry replication workflows, aligning with **Agile + CI/CD**, and reducing refresh times to accelerate research iterations.
- Generated statistically preserved datasets, boosting reproducibility, compliance, and large-scale ML experimentation traceability.

3.The Sparks Foundation — March 2022 – April 2022
Web Development Intern — *Bangalore, India*

- Developed a full-stack **data management platform** using **HTML, CSS, JavaScript**, supporting secure and scalable data workflows.
- Optimized backend **MySQL schemas** for efficient queries and structured storage, enabling reproducibility and analytics-driven experimentation.
- Deployed the platform on **AWS with Docker**, integrating **CI/CD pipelines** to ensure scalability, reliability, and cloud-native performance.

Projects

1.Tool Research AI-Agent |*LangChain, LangGraph, LangSmith, Python, Firecrawl*

- Created an research agent integrating LLMs with web scraping to automate tool evaluation, reducing manual research effort.
- Extracted structured insights from APIs and technical docs, showcasing ability to integrate LLM reasoning into applied decision systems for scalable AI-driven data platforms.

2.Transfer Learning for Pneumonia Detection | *Python, Keras, TensorFlow, ImageNet*

- Implemented **transfer learning models** (ResNet, DenseNet, InceptionV3), achieving **96% accuracy, 90% precision, 88% recall, and 89% F1-score** on medical imaging datasets with augmentation, cross-validation, and hyperparameter tuning.
- Built reproducible ML workflows with **explainability techniques (SHAP/LIME)**, ensuring transparency, supporting clinical research, and demonstrating practical ML deployment in healthcare use cases.

Achievements

Google Cloud Badges: Prompt Design in Vertex AI (Apr 2025) • Build Real-World AI Applications with Gemini and Imagen (Apr 2025) • Develop GenAI Apps with Gemini and Streamlit (Apr 2025) • Inspect Rich Documents with Gemini Multimodality and RAG (Apr 2025) • Explore Generative AI with Vertex AI Gemini API (Apr 2025).