Signature

MANI KANT PANDEY

+91-6307169317 ◇ Greater Noida, India
pmanikant2@gmail.com ◇ linkedin.com/in/mani-kant-pandey ◇ github.com/manikantpandey

SKILLS

Languages & Libraries	Python, Rust, CUDA, TensorFlow, PyTorch, Keras, Scikit-learn, LangChain
AI / GenAI	LLM Fine-Tuning, NLP, Deep Learning, GANs, RAG, Vector DB, MCP, Hugging Face
MLOps & DevOps	FastAPI, Jenkins, CI/CD, Docker, Kubernetes, Experiment Tracking, Model Versioning
Cloud	Azure, Google Cloud Platform
Other Tools	MySQL, Pandas, NumPy, Matplotlib, Selenium, BeautifulSoup, Playwright,

EXPERIENCE

Machine Learning Engineer Dec 2024 – Present
Nuster AI *Remote*

- Engineered an autonomous, agentic AI pipeline using LLMs and prompt engineering, reducing decision-latency by 35%.
- Designed backend infrastructure to deploy AI-driven fintech services with latency <200 ms and 99.9% uptime.
- Migrated core backend from Python to Rust, improving execution speed, concurrency handling, and memory safety.
- Built scalable MLOps workflows using Jenkins for automated retraining, testing, and deployment on cloud platforms.
- Integrated multi-agent orchestration for financial data analysis, fraud detection, and personalization.
- Optimized inference pipelines, cutting GPU costs by 20% through batch processing and model quantization.

Machine Learning Intern Jul 2024 – Nov 2024
Jarviscalling.AI *Remote*

- Developed a conversational NLP bot powered by LLMs, reducing average customer response time by 25%.
- Automated large-scale data extraction from 200K+ records using Selenium, improving training efficiency by 40%.
- Implemented a QA system using FastAPI and LLM frameworks, boosting query accuracy by 30%.
- Fine-tuned transformer-based models for customer intent classification, achieving 92% accuracy.
- Designed a lightweight caching mechanism for model responses, reducing API latency by 18%.

Machine Learning Intern Jul 2022 – Aug 2022
Omnipresent Robot Tech *Onsite*

- Contributed to an autonomous warehousing drone system, boosting inventory throughput by 40%.
- Optimized point cloud mapping algorithms, reducing drone collision risk by 15%.
- Applied deep RL techniques to enhance real-time object detection accuracy by 25%.

PROJECTS

Customer Satisfaction Predictor. Developed an end-to-end pipeline using ZenML and XGBoost, boosting prediction accuracy from baseline by 30% (to 85%).

Contextual Chatbot with FAISS & BERT. Built a retrieval-based conversational bot using BERT embeddings and FAISS, enhancing search speed by 35%.

EDUCATION

Gautam Buddha University 2021 – 2025
B.Tech in Computer Science & Engineering, Specialization in Machine Learning Greater Noida, India

Allahabad Public School 2011 – 2021
Primary and Secondary Education, Prayagraj, India

Initial Signature

Initial Signature