Neel Patel

neelp265@gmail.com | (267) 403-9524 | linkedin.com/in/neelpatel265 | github.com/gh-neelpatel

Education

Drexel University, M.S. in Artificial Intelligence & Machine Learning, GPA: 3.97/4.00 Expected June 2026
Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, Probabilistic Modeling, Algorithms, DSA
Dharmsinh Desai University, B.S. in Electronics & Communications Engineering, GPA: 3.60/4.00 May 2023
Coursework: Computer Architecture, Parallel Computing, Microprocessors, DSP, Probability & Statistics

Skills

ML, Deep Learning & Modeling: PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, CNNs, RNNs, GRUs, LSTMs, Transformers, Time-Series Modeling, Sequence Modeling, Probability, Statistics, Model Evaluation
MLOps & Model Lifecycle: MLflow, Model Versioning, Experiment Tracking, CI/CD for ML, KPI Regression Tracking, Drift Monitoring, Model Monitoring, Reproducible Pipelines, DeepEval, Langfuse
Cloud, Data & Infrastructure: AWS (SageMaker, EC2, S3, Lambda), Databricks, PySpark, PostgreSQL, Docker, Kubernetes, OpenShift, Helm, Terraform, Jenkins, GitHub Actions, Linux, Slurm
Programming & Systems: Python, C/C++, SQL, Java, Bash, REST APIs, Microservices, OOP Design, Agile/Scrum
AI Systems: LangChain, LangGraph, Pinecone, RAG, GPT-4, LLaMA 3, HuggingFace, OpenAI API, AI Agents, MCP

Experience

Machine Learning Operations Engineer, Boehringer Ingelheim – Ridgefield, CT September 2025 – Present
- Built end-to-end ML retraining pipelines across AWS, OpenShift, Jenkins, and Slurm GPUs, automating data ingestion, feature generation, training, evaluation, and deployment for large-scale model workloads
- Deployed containerized ML workloads with Docker, Helm, Kubernetes, and OpenShift using Terraform IaC, cutting deployment time by 25% while improving reproducibility and release governance
- Implemented model evaluation and release monitoring with MLflow, DeepEval, and Langfuse, tracking KPI regressions, data drift, model quality, and reliability metrics across production releases
- Architected production ML microservices with PostgreSQL, REST APIs, and governed data workflows, enabling secure model integration, lifecycle automation, and reliable deployment for regulated teams

Software Engineer, Matrix Comsec Pvt. Ltd. – Vadodara, India January 2023 – April 2024
- Built production Python and C/C++ services for Linux-based telecom systems across 15+ repositories, improving real-time signal processing reliability and integrated release execution
- Optimized Jenkins CI/CD pipelines through bottleneck profiling, parallel execution, and automated validation, increasing throughput by 20% and reducing deployment time from 60 to 36 minutes
- Implemented DevSecOps gates with Parasoft, Nessus, and OWASP across microservice repositories, reducing vulnerability resolution time by 40% and preventing critical production defects
- Built real-time monitoring and alerting for telecom release pipelines, surfacing security metrics, failed validations, and regression signals before issues reached production users

Projects

GlassBox, Agentic AI Evaluation Platform | Python, React 19, TypeScript, MCP, OpenAI API, GPT-4
- Engineered full-stack AI evaluation platform processing 17,000+ agent execution events, tracking reasoning traces, tool outcomes, KPI deltas, and behavior metrics across 36 AI behavior types
- Built MCP-integrated compliance reports for HIPAA, GDPR, SOC2, and AI Act with multi-dataset delta analysis and a custom Python parser, winning 1st Place at PhillyCodefest 2026 ($3,000 prize)

ML Fine-Tuning and Optimization Pipeline | Python, PyTorch, TensorFlow, Keras, HuggingFace, LoRA, CUDA
- Built distributed training pipeline for LLaMA 3 using PyTorch, LoRA, and PEFT, configuring multi-GPU training, mixed precision, gradient accumulation, and Linux GPU infrastructure
- Optimized inference with FP16/INT8 quantization experiments and latency profiling, measuring perplexity, accuracy, and serving tradeoffs across baseline and compressed variants for constrained deployment