PIJUSH PAL

AI Engineer | Semantic Analysis Pipelines | Prompt Engineering Strategies
(+91) 8250 727 328 • pijushpl2023@gmail.com • LinkedIn | GitHub • Bangalore, INDIA

Summary

AI Engineer with over 5 years of experience in building AI applications and advanced LLM systems, expert in retrieval-augmented generation (RAG) and deep learning techniques. My key achievements include architecting multi-agent LLM workflows that improved anomaly detection accuracy by 30% and designed end-to-end RAG pipelines that increased contextual retrieval precision by 60%. Seeking an AI Engineer position at Netlink Software America, where I bring my AI application and deep learning skills to support your mission of protecting users from modern threats through cutting-edge scam detection solutions.

Key Achievements

Anomaly Detection Enhancement

Improved anomaly detection accuracy by 30% using model fine-tuning techniques in AI workflows.

Contextual Retrieval Precision

Increased contextual retrieval precision by 60% through design of retrieval-augmented generation pipelines.

Response Latency Reduction

Reduced response latency by 35% with advanced model routing using retrieval-augmented generation methods.

Experience

Netlink Software Group America — Bangalore, IN

Senior AI Engineer — 10/2025

- Built production-grade LLM and Argentic AI systems for enterprise intelligence, anomaly detection, and high-stakes decision workflows
- Architected multi-agent LLM workflows which improved anomaly detection accuracy by 30% while reducing manual investigation time by 40%
- Designed end-to-end RAG pipelines, increasing contextual retrieval precision by 60% and reducing hallucination rate by 20%
- Implemented prompt optimization and model routing, resulting in reduced response latency by 35% and inference cost by 25%
- Built observability framework tracking model drift, improving model reliability in high-stakes workflows

Wekare.INC — Texas, US

Gen-AI Engineer — 03/2025 - 10/2025

- Developed production-grade LLM systems for high-stakes decision automation in healthcare workflows
- Designed a client-server agentic framework, which improved multistep reasoning accuracy by 25%
- Deployed scalable AI services using AWS SageMaker, reducing response latency by 50%
- Built evaluation pipelines to monitor model drift and reduced high-risk output incidents in production
- Implemented advanced NLP pipelines for improved operational efficiency

Mangoes.ai — New York, US

AI/ML Engineer — 03/2024 - 02/2025

- Developed RAG and multimodal AI systems for contextual search and knowledge intelligence
- Architected advanced pipelines integrating structured + unstructured data, improving retrieval relevance by 30%
- Fine-tuned LLMs for improved accuracy and domain alignment
- Built semantic reranking pipelines to reduce inference latency by 45%
- Standardized and optimized NLP model deployment for improved system stability

University of New Haven — Connecticut, US

ML Engineer — 01/2023 - 07/2023

- Developed predictive decision systems for graduate admissions
- Built ML-based decision system, reducing review time from 20 to 12 minutes (40% efficiency gain)
- Led 15-member ML team for end-to-end predictive system development, achieving 75–80% predictive accuracy

Ineuron.ai — Bangalore, IN

ML Engineer — 01/2021 - 09/2022

- Built ML/DL systems for business analytics and predictive modeling
- Implemented supervised models and neural networks, enhancing prediction reliability by 35–40%
- Applied clustering and time-series modeling techniques for better accuracy

Freelance (Self-Employed) — Kolkata, IN

Full Stack AI — 08/2017 - 08/2019

- Delivered full-stack development and early-stage ML solutions for small business clients
- Designed and developed end-to-end applications enhancing operational efficiency
- Diagnosed and resolved issues, reducing downtime for client environments
- Implemented predictive modeling to enable data-driven decision-making

Experience

Welspot.ai — Florida, US

AI Intern — 01/2024 - 05/2024

- Achieved 30% accuracy increase while reducing cloud costs by 45% through optimized ML model performance.
- Built ML pipelines on AWS Sage Maker, reducing model deployment time by ~30%.
- Implemented A/B testing frameworks for improved predictive accuracy.
- Automated CI/CD workflows, increasing release reliability.

Education

Duke University (Online) — 01/2025 - 05/2026

Master's Degree

University of New Haven — 01/2023 - 05/2024

Master's Degree

WBUT — 01/2014 - 05/2018

Bachelor's Degree

Skills

Gmail · GitHub · AWS · Azure · Lora · AWS SageMaker · Python · PyTorch · Sagemaker · AWS Lambda · S3 · EC2 · Azure Machine Learning · Docker · Kubernetes · Caching · NLP · Streamlit · XGBoost · LLM · RLHF · AB Testing · Cybersecurity · Deep Learning · TensorFlow · Kafka · Kinesis · Adaptability · Teamwork · Critical Thinking · Leadership · Coaching