

MANTHAN C

Senior Senior AI/ML Engineer

SUMMARY

- Senior AI/ML Engineer with 5+ years deploying generative AI and ML systems using GPT-4, Claude 3, LLaMA, and LangChain.
- Expertise in Python backend development (FastAPI, Django) serving LLMs and ML models with scalable cloud deployment on AWS, Azure, GCP.
- Built enterprise RAG platforms with hybrid retrieval (FAISS, Neo4j) and multi-agent orchestration reducing hallucinations and improving LLM output quality.
- Developed AI pipelines for sales enablement, GDPR compliance, semantic multimodal search, and predictive modeling using PyTorch, TensorFlow, LightGBM, and CatBoost.

TECHNICAL SKILLS

Main Technical Skills	Python (5 yr.), FastAPI (5 yr.), Django (5 yr.), LangChain (5 yr.), PyTorch (5 yr.)
AI Tools & Assistants	Claude (3 yr.)
Programming Languages	Python (5 yr.)
AI & Machine Learning	AWS SageMaker (3 yr.), CatBoost (3 yr.), Claude (3 yr.), Computer Vision (5 yr.), Faiss (5 yr.), GPT (3 yr.), Hugging Face (3 yr.), LangChain (5 yr.), LangGraph (3 yr.), LLaMA (3 yr.), LlamaIndex (3 yr.), Mistral AI (3 yr.), NLP (5 yr.), NumPy (1 yr.), OpenAI (3 yr.), OpenCV (1 yr.), Prompt Engineering (3 yr.), PyTorch (5 yr.), RAG (3 yr.)
Python Frameworks	Django (5 yr.), FastAPI (5 yr.), Flask (5 yr.)
Java Libraries and Tools	JSON Schema (1 yr.)
JavaScript Frameworks	Next.js (1 yr.)
Python Libraries and Tools	NumPy (1 yr.), Pandas (1 yr.), PyQt (1 yr.), PyTorch (5 yr.), Robot Framework (1 yr.)
Data Analysis and Visualization Technologies	Apache Airflow (3 yr.), ARIMA (3 yr.), LightGBM (3 yr.), Pandas (1 yr.)
Security	Auth0 (1 yr.)
Databases & Management Systems / ORM	AWS DynamoDB (3 yr.), ChromaDB (3 yr.), MySQL (3 yr.), Neo4j (3 yr.), PostgreSQL (3 yr.), Redis (3 yr.)
Cloud Platforms, Services & Computing	Azure ML (3 yr.), GCP (3 yr.)
Amazon Web Services	

	AWS Bedrock (3 yr.), AWS CloudWatch (3 yr.), AWS DynamoDB (3 yr.), AWS EC2 (3 yr.), AWS Lambda (3 yr.), AWS S3 (3 yr.), AWS SageMaker (3 yr.)
Azure Cloud Services	Azure AD hybrid identity (3 yr.), Azure ML (3 yr.)
Industry Domain Experience	Banking (3 yr.), Embedded (3 yr.)
Message/Queue/Task Brokers	Celery (3 yr.)
Deployment, CI/CD & Administration	CI/CD (5 yr.), Jenkins (3 yr.)
SDK / API and Integrations	FastAPI (5 yr.), RESTful API (5 yr.)
Version Control	Github Actions (3 yr.)
Web/App Servers, Middleware	gUnicorn (3 yr.), Nginx (3 yr.)
UI/UX/Wireframing	UED (1 yr.)
Mail / Network Protocols / Data transfer	WebSockets (5 yr.)
Other Technical Skills	async (5 yr.), Fine-Tuning SonarQube (3 yr.), Illustration and icons (3 yr.), Milvus (3 yr.), SHAP (3 yr.), Weaviate (3 yr.)

WORK EXPERIENCE

Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (Enterprise RAG Knowledge Assistant)

Duration: Feb 2021-Present

Summary:

- Built an enterprise Retrieval-Augmented Generation (RAG) platform that answers questions from company knowledge using hybrid retrieval combining vector search and graph traversal
- The platform leverages Neo4j to model relationships between documents and entities for enhanced context and uses multiple open-source LLMs with multi-stage reranking to improve answer quality

Responsibilities:

- Designed and implemented hybrid retrieval system using vector databases and graph databases.
- Developed routing mechanisms across 50+ open-source LLMs and multi-stage reranking with cross-encoders.
- Implemented hallucination reduction techniques by enforcing structured output.
- Built authentication, monitoring, and evaluation suites to benchmark LLM quality.

Technologies: Python, FastAPI, LangChain, Neo4j, Hugging Face, FAISS, Runpod



Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (PersonaGrid: AI Sales Enablement Platform)

Duration: Feb 2021-Present

Summary:

- Developed an AI platform for sales call simulation and pre-call planning that supports multi-tenant data isolation and structured AI output
- The platform enables sales teams to prepare effectively by simulating calls and reasoning through multi-step AI pipelines

Responsibilities:

- Designed multi-step reasoning pipelines using LangChain chaining.
- Implemented multi-tenant vector database architecture to ensure data separation per organization.
- Enforced JSON schema to maintain structured and compliant AI outputs.
- Managed background task execution with Celery and Redis and automated organizational setup through CI/CD pipelines.

Technologies: Python, FastAPI, LangChain, OpenAI, Celery, Redis, Next.js, AWS

Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (AI-Powered Data Protection Compliance Portal)

Duration: Feb 2021-Present

Summary:

- Built a GDPR compliance portal that automates data protection queries and risk audits using AI
- The portal assists compliance teams by providing AI-assisted PII redaction, severity scoring, and automated responses to reduce manual workload

Responsibilities:

- Developed AI pipelines for PII redaction, severity scoring, and automated query responses using Claude 3.
- Implemented audit scoring and visual reporting features for compliance teams.
- Integrated authentication with Auth0 and deployed on AWS infrastructure.

Technologies: Python, FastAPI, LangChain, ChromaDB, Claude 3, Auth0, AWS

Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (Semantic Text & Visual Search Engine)

Duration: Feb 2021-Present

Summary:

- Developed a semantic search engine supporting both text and image queries using embedding-based retrieval
- The system enables multimodal search capabilities and scalable model deployment for efficient and accurate search results



Responsibilities:

- Built embedding-based retrieval system with FAISS indexing.
- Implemented multimodal search supporting text and image queries.
- Developed image similarity pipelines with experiment tracking for model evaluation.
- Deployed and managed models on AWS SageMaker with monitoring and scalable inference.

Technologies: Python, TensorFlow, Hugging Face, FAISS, AWS SageMaker, S3, DynamoDB

Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (Horse Race Outcome Prediction System)

Duration: Feb 2021-Present

Summary: Built a machine learning system to predict horse race outcomes including final position and return on investment (ROI) using ensemble models trained on large historical datasets.

Responsibilities:

- Developed stacked ensemble models using LightGBM and CatBoost.
- Designed feature engineering and validation pipelines to ensure data quality and reproducibility.
- Conducted repeated experiments and model tuning to improve prediction accuracy.

Technologies: Python, scikit-learn, LightGBM, CatBoost, Pandas, NumPy

Senior AI/ML Engineer - Cyber Infrastructure (CIS) Pvt. Ltd. (OCT Medical Image Segmentation)

Duration: Feb 2021-Present

Summary:

- Developed a desktop application for segmenting optical coherence tomography (OCT) medical eye scan images using deep learning
- The application assists medical professionals by providing accurate segmentation and measurement extraction

Responsibilities:

- Developed a custom U-Net model for mask prediction on OCT images.
- Used OpenCV for image masking and measurement extraction.
- Applied transfer learning to improve classification accuracy.
- Wrapped the model in a PyQt interface for desktop application usability.

Technologies: Python, TensorFlow, OpenCV, U-Net, Roboflow, PyQt

EDUCATION

- **Bachelor of Technology (B.Tech), Computer Science**
Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV)



CERTIFICATION

- **TensorFlow Developer Certificate**
Udemy
- **Deep Neural Networks with PyTorch**
Coursera
- **Machine Learning Bootcamp**
Udemy
- **Computer Vision with OpenCV and Deep Learning**
Udemy

