

PROFESSIONAL SUMMARY

AI Engineer with 1+ years of experience building production ML pipelines, RAG systems, and data representation workflows using Python, Pandas, NumPy, and Apache Spark. Hands-on with embedding workflows, vector storage, knowledge graph construction, and retrieval-augmented generation. Strong foundation in supervised/unsupervised ML, feature engineering, structured debugging, and Git-based CI/CD — with a track record of delivering clean, testable, well-documented ML systems end-to-end.

WORK EXPERIENCE

Palcode.ai

AI Engineer

Indore, India

June 2025 - Sept 2025

- Deployed RAG-style NLP workflows on AWS EC2 — implementing document chunking, metadata tagging, prompt/context templates, and OpenAI API-based retrieval pipelines for post-call summarization and entity extraction.
- Built fault-tolerant event-driven batch pipelines on AWS Lambda and EventBridge with structured logging, reproducible configs, and incident triage support — following Git workflows and CI checks throughout.
- Maintained evaluation sets for LLM outputs, performing retrieval quality checks and answer quality regressions — documenting pipeline SOPs and dataset behavior across all production service layers.

Superteams.ai

AI Developer

Indore, India

Aug 2024 - May 2025

- Designed scalable feature engineering pipelines using Apache Spark (PySpark) and NVIDIA RAPIDS on Linux — implementing aggregations, window functions, and data validation to cut execution time by 50%.
 - Built RAG-based document parsing workflows with Django REST APIs and MongoDB — applying data profiling, joins, and anomaly detection with Graylog monitoring to improve system efficiency by 80%.
 - Implemented a Flux.1-based image generation ML system using PyTorch fine-tuning — logging experiments, tracking model outcomes, and maintaining reproducible training configs under CI/CD pipelines.
-

PROJECT WORK

Knowledge Graph — Embedding Pipeline with Vector & Graph Storage

- Built an end-to-end data representation pipeline using GPT-4o and LlamaIndex — extracting 50+ semantic relationships from unstructured text, constructing canonical entity dictionaries, and indexing embeddings into Neo4j graph database with vector retrieval for optimized query resolution.
- Applied chunking strategies, metadata tagging, and hybrid retrieval concepts across ingestion, transformation, and retrieval stages — reducing schema creation time by 65% through systematic optimization.

SkillSprint-AI — RAG-Enabled ML System with Evaluation Pipeline

- Developed a FastAPI backend integrating LLM APIs with prompt/context templates, grounding rules, and structured evaluation sets — running offline retrieval quality and answer quality regressions across releases.
- Implemented reproducible configs, unit tests, and Git-based CI checks; maintained dataset cards and feature notes for all pipeline components with full deployment documentation.

Food Label Scanner — ML Inference Pipeline with Error Analysis

- Built a modular FastAPI ML inference pipeline with layered OCR, Vision AI, and model-based parsing — performing systematic error analysis and data leakage checks that improved extraction accuracy from 60% to 90%.
- Wrote clean, testable Python code with unit tests across all service layers; documented API contracts, pipeline SOPs, and model behavior for traceability and long-term maintainability.

CINEXUS — Distributed ML-Ready Backend with Feature Store Concepts

- Architected a distributed Python/Django backend on AWS EC2 with PostgreSQL, implementing batch data aggregations, real-time scoring APIs, and analytics dashboards with reproducible deployment configs for 1,000+ concurrent users.
-

CERTIFICATIONS

- Oracle Cloud Infrastructure 2025 — Certified Generative AI Professional
 - TCS CodeVita Season 12 — Global Rank 1470
-

EDUCATION

Bachelor of Technology in Information Technology

MediCaps University CGPA: 8.14

Indore, India

Aug 2021 - July 2025

TECHNICAL PROFICIENCY

Languages: Python, SQL, Java, JavaScript

ML & Data: Pandas, NumPy, PySpark, PyTorch, Supervised/Unsupervised Learning, Embeddings, Feature Engineering, Data Validation

RAG & Retrieval: LlamaIndex, RAG Pipelines, Vector Indexing, Chunking Strategies, Prompt Engineering, OpenAI API, Neo4j

MLOps & Engineering: Experiment Tracking, Reproducible Configs, Unit Testing, Structured Debugging, Logging, CI/CD, Git

Cloud & Infra: AWS (EC2, Lambda, EventBridge, S3, RDS), MongoDB, PostgreSQL, Docker, Kubernetes, Linux, Apache Spark

Key Skills: Data Representation, Knowledge Graphs, Entity Extraction, Pipeline SOPs, Dataset Documentation, Agile