

AMANDEEP

Gurgaon, India | +91 9810912788 | amanjyani98@gmail.com | iamamandeep.in | linkedin.com/in/amanjyani9 | github.com/amanjyani9

PROFILE SUMMARY

AI/ML Engineer with 4+ years of experience building, training, and deploying machine learning models and agentic AI systems in production environments. Expertise in Large Language Models (GPT-4, Gemini) and agentic workflows (LangGraph) for text, conversational, and automation applications, with strong proficiency in Python (pandas, NumPy, scikit-learn). Proven ability to design end-to-end ML and agentic pipelines including data cleaning, feature engineering, model optimization, and evaluation. Excellent analytical thinking and cross-functional collaboration skills.

TECHNICAL SKILLS

ML / AI: Agentic AI (LangGraph), LLMs (GPT-4, Gemini), NLP, Classification, Regression, Model Optimization, OCR (Tesseract)

Languages: Python, SQL, PySpark

Libraries: Pandas, NumPy, Scikit-learn, TensorFlow, OpenAI API, LangChain, LangGraph, Transformers

ML Ops: Model Training, Feature Engineering, Data Cleaning, Model Evaluation, A/B Testing

Data / ETL: Databricks, Apache Spark, ETL Pipeline Design, Data Modeling

Cloud: Azure (AKS, Key Vault), Datadog, BigQuery

Tools: Jupyter, VS Code, Power BI, Postman, Git

SDLC: Agile (Scrum/Rally), Cross-functional Collaboration, Documentation

PROFESSIONAL EXPERIENCE

Optum Global Solutions (UnitedHealth Group) — Gurgaon, India

AI/ML Engineer

Jul 2026 – Present

- **Designed and deployed an agentic transcript-analysis system using LangGraph and the OpenAI API**, automating end-to-end review and intent-tagging that previously required manual analysis — cutting human effort from 8 hours to 1 hour of validation per batch.
- Analyzed conversational bot transcripts using multiple clustering schemes and LLMs to recognize caller intent, boosting intent-recognition accuracy from 80% to 88% and lifting authentication success rate from 85% to 90%.
- Partnered with Product and Marketing teams to visualize transcript analytics and surface authentication-gap and call-driver insights, directly informing roadmap and messaging decisions that improved customer self-service outcomes.
- Built and optimized LLM-based applications (GPT-4, Gemini) for text extraction, conversational AI, and automated summarization — engineering prompts for consistent structured JSON output in production.
- Built Gemini model variant classification system using statistical analysis of token patterns to categorize LLM requests into pricing tiers — demonstrating strong analytical thinking.

Software Engineer

Jul 2023 – Jun 2026

- Trained and deployed OCR model (Tesseract) with pre-processing pipelines (deskew, binarization, noise removal) and model evaluation metrics (precision, recall, F1) for production quality assurance.
- Collaborated cross-functionally with Finance, Product, and Engineering teams to translate business requirements into ML-driven solutions with clear documentation and stakeholder presentations.
- Designed feature engineering pipelines transforming raw log data into aggregated metrics across multiple dimensions, applying data cleaning, outlier detection, and normalization techniques.
- Developed async parallel ML inference pipelines (Python, pandas, NumPy) chunking documents into token-limit blocks for production-scale LLM processing with error handling and retry logic.
- Created user behavior analytics models aggregating session data, identifying usage patterns, and building segmentation features for cross-functional business intelligence reporting.
- Implemented data cleaning and validation frameworks comparing API-sourced metrics against database records, applying scikit-learn anomaly detection to flag inconsistencies.

Training Development Program (TDP II)

Jul 2022 – Jul 2023

- Developed a ChatGPT-powered automation for text and conversational applications, training team members on prompt design and LLM best practices.
- Built data cleaning and model evaluation workflows ensuring accuracy and reliability in ML outputs.

KEY PROJECTS

Agentic Transcript-Analysis System (LangGraph, OpenAI API)	2026
- Built a multi-step agentic pipeline in LangGraph to ingest, classify, and summarize conversational transcripts, orchestrating tool calls and structured outputs across graph nodes. - Reduced required human review from 8 hours to 1 hour of validation per batch by handing routine analysis to the agent and reserving human effort for edge cases.	
LLM-Powered Text & Conversational Applications	2025 – 2026
- Built production LLM applications using GPT-4 and Gemini for text classification, structured extraction, and conversational workflows with model evaluation and optimization. - Designed variant classification ML system categorizing requests into 6 tiers using feature engineering on token counts, context length, and usage patterns.	
ML Inference API (GPT-3.5/4, Python)	Nov – Dec 2024
Built async ML inference service with pandas/NumPy preprocessing, prompt optimization, and model evaluation — achieving 7x throughput improvement via parallel processing.	
OCR Model Training & Deployment	2024
Trained and deployed Tesseract OCR model with custom pre-processing pipeline, feature engineering for document layout analysis, and production model evaluation (F1 > 0.92).	

EDUCATION

Malviya National Institute of Technology	Sep 2020 – Jul 2022
<i>M.Tech, Computer Science GPA: 8.28/10 — Jaipur, India</i>	
Amity University	Jul 2015 – Jul 2019
<i>B.Tech, Computer Science GPA: 8.14/10 — Gurgaon, India</i>	

CERTIFICATIONS

- Microsoft Azure Fundamentals — PluralSight (Jun 2024)
- Python Essentials for MLOps — Duke University, Coursera (Feb 2024)
- Machine Learning with Python — IBM, Coursera (Nov 2022)