

ANKIT MANDUSIA

AI Engineer

mandusiaankit.005@gmail.com
linkedin.com/in/ankit-mandusia-8817a916b
https://ankitmandusia.github.io/
+91-7726856379
Jaipur

AI Engineer driving real-world AI solutions in chat, voice, and vision, with expertise in generative AI and multimodal systems. Experienced in RAG pipelines, knowledge graphs, and intelligent document automation to deliver business impact.

Work Experience

Project Associate I

Nov 2024 - Present

Malaviya National Institute of Technology | Jaipur, Rajasthan

- Designed and implemented data-driven workflows for automated battery testing, integrating realtime sensor data using Arduino and Raspberry Pi.
- Applied machine learning techniques and statistics for analyzing battery performance datasets, identifying patterns in charging/discharging cycles, and improving predictive modeling of cell behavior.
- Developed ETL pipelines for DOBOT CR3 and MG400 test benches, producing clean datasets for analytics.
- Video Demonstration – Robotics Automation Setup - <https://youtu.be/4Aygwmuk-TU>

Freelance AI Training Expert

Oct 2024 - Mar 2025

Outlier | Remote, India

- Contributed to RLHF and SRT projects (Hopper, Cypher Evals, Matcha Stem, Pref Ranking, Extensions V2 Log), enhancing generative AI training through code evaluation and task curation.

Projects

SEC Filing Q&A with GraphRAG

May 2025 - Aug 2025

- Architected and developed a state-of-the-art multi-agent RAG system using LangChain and a Mistral-7B model to enable advanced, context-aware Q&A and analysis of dense financial documents (SEC 10-K filings).
- Engineered a "Natural Language to Query" feature, enabling an AI agent to translate user requests into reliable, on-the-fly data visualizations with Plotly.
- Developed a full-stack analysis tool with an interactive Streamlit UI, leveraging Qdrant for vector search and Pandas for robust data ingestion.
- Technologies: LangChain, LlamaIndex, Hugging Face, Qdrant, Transformers, Streamlit

Knowledge Assistant with Multimodal RAG

Jun 2025 - Aug 2025

- Architected a complete RAG pipeline using LangChain and a Pinecone vector database, enabling accurate, context-aware conversations.
- Deployed and optimized multiple open-source LLMs for local inference, using 4-bit quantization to run large models on resource-constrained GPUs.
- Built the full-stack application with a FastAPI backend, a robust PyMuPDF data ingestion pipeline, and a dynamic JavaScript frontend.
- Technologies: LangChain, FastAPI, Pinecone, RAG, Llama 2, 4-bit Quantization

- Built a voice-interactive assistant using Gemini 2.0 API to provide fact-based responses from audio or text queries.
- Implemented real-time speech-to-text and text-to-speech pipelines for seamless two-way voice interaction using , gTTS, and Pydub, reducing response latency to under 1.5 seconds and improving interaction efficiency by 30%.

Automatidata - ML for Taxi Revenue Optimization

Jan 2024 - Apr 2024

- Developed and deployed a multiple linear regression model achieving an R² of 0.84, improving fare estimation accuracy by 28% compared to baseline methods
- Conducted detailed inspection , EDA, A/B Testing, and feature engineering on 100K+ trip records

Core Skills

Python, R, C, SQL, Regression analysis, Excel/Spreadsheet, GCP, NLP, Azure, Power BI, Pandas, Quantitative Analysis, AWS, Predictive Models, LLM, RAG, Langchain

Education

Dayalbagh Educational Institute, Agra

Aug 2022 - Jun 2024

Master of Technology Computer Science
GPA: 7.36

BK Birla Institute of Engineering and Technology, Pilani

Aug 2017 - Oct 2021

Bachelor of Technology Information Technology
GPA: 7.20

Certificates

Data Analytics Professional Certificate

Jul 2023

Google

Advanced Data Analytics Professional Certificate

Feb 2024

Google

AzureML Model for Data Scientist

Feb 2023

Microsoft

Qiskit Global Summer School 2025 - Quantum Excellence

Aug 2025

IBM

Publications

Brain tumor segmentation using deep learning-based mri analysis

Aug 2024

International Journal of Engineering Research Technology (IJERT)