Anuj Patel
Data Scientist | Machine Learning Engineer | AI Engineer
Arlington, Texas | +1 682-217-2520 | anuj.patel.29dec@gmail.com | LinkedIn | Portfolio | Medium | GitHub
MS in Computer Science (AI/Cloud), University of Texas at Arlington

---

I'm a Data Scientist with a strong background in machine learning, predictive analytics, NLP, and deep learning. I enjoy building intelligent and scalable data solutions that automate real-world processes and solve meaningful problems. Over the years, I've worked on complete ML pipelines—from gathering and cleaning data to developing models, deploying them, and keeping them running smoothly in production.

I work comfortably across Python, Scikit-learn, TensorFlow, PyTorch, SQL, and cloud platforms, and have hands-on experience with LLMs and OCR/NLP systems. With a blend of Data Science, MLOps, and Software Engineering skills, I focus on creating reliable, production-ready solutions that genuinely move the needle for businesses.

- **University of Texas at Arlington — Master of Science in Computer Science** 2025–2027
  Specialisation : Artificial Intelligence & Cloud Computing
  Coursework: Artificial Intelligence, Machine Learning, Cloud Computing, Data Structures
- **PIEMR, Indore — B.Tech in Computer Science & Engineering** 2021–2025
  Coursework: OOP, Databases, Software Design Patterns, Data Structures

EXPERIENCE

Software Developer – RinaySoft, India June 2024 – June 2025

- Designed machine-learning-powered automation systems, integrating OCR + LLMs for structured data extraction.
- Improved **database performance by 30%** through schema optimisation and query tuning.
- Developed RESTful APIs enabling analytics, reporting, and automated data pipelines.
- Built backend workflows for **data ingestion, preprocessing, and ML-driven classification**.
- Collaborated on system architecture and cross-functional review cycles for scalable data systems.

Python Intern – Renuka Softec, India April 2024 – May 2024

- Rebuilt a legacy application, reducing system errors by **20%** and increasing data reliability.
- Implemented automated data processing workflows, reducing manual work by **40%**.
- Optimised SQL queries and improved database efficiency, enabling smoother analytical operations.

Placement Cell — Head Coordinator 2022 - 2025

- Coordinated with multiple companies to conduct campus recruitment drives.
- Led and managed the entire placement process, from scheduling to final interviews.
- Directed a team of volunteers and ensured seamless communication between companies and students.

DATA SCIENCE PROJECTS

---

1. Customer Churn Prediction System || ML Classification || Predictive Analytics || LINK

- Developed a churn prediction model using supervised ML algorithms (Logistic Regression, Random Forest, XGBoost).
- Performed feature engineering, exploratory analysis, and hyperparameter tuning.
- Provided actionable insights for retention strategies based on model outputs.

2. Automated Bill Extraction System (OCR + LLM) || Data Ingestion || NLP || LINK

- Created a Streamlit application to extract structured fields (date, amount, vendor) from scanned bills.
- Improved OCR accuracy by **30%** using preprocessing and Google Vision API.
- Used LLM-based entity extraction (Hugging Face), achieving **85% classification accuracy**.

3. Twitter Sentiment Analysis NLP || Text Classification || Flask Deployment || LINK

- Built a sentiment classification model (~80% accuracy) using NLP preprocessing and ML algorithms.
- Developed a web interface for real-time analysis via Flask API endpoints.
- Conducted cleaning, tokenization, stop-word removal, and vectorization.

4. RAG Model for Information Retrieval || Vector Search || Embeddings || LangChain || LINK

- Built a hybrid search system combining semantic embeddings + contextual LLM responses.
- Implemented ChromaDB vector store for dense embedding search.
- Designed an evaluation workflow for retrieval quality and output coherence.

5. Multi-Agent Orchestrator System with Standardized A2A Protocol || LINK

- Built a multi-agent orchestration framework enabling autonomous agent-to-agent communication using a standardized A2A protocol.
- Designed a centralized orchestrator for task routing, context sharing, and structured message passing between specialized agents.
- Coordinated interactions between multiple users' agents by exchanging availability data and computing their mutual free time for the main scheduling agent.

RESEARCH PUBLICATION

- **A Study on the Potential of AI in the Healthcare Sector ||** Metszet Journal 2024
  metzet 4303 R.pdf

STRENGTHS

- Excellent communication skills with the ability to explain complex technical concepts clearly
- Fast learner with the ability to quickly adapt to new tools, technologies, and methodologies.
- Proven leadership skills through managing teams, coordinating events, and driving initiatives
- Research-oriented thinking with a strong inclination toward experimentation and continuous improvement