

PULKIT MANGAL



IIT Madras
B.S. Degree

8209446456 | pulkitmangal09@gmail.com | [linkedin.com/in/its-pulkit](https://www.linkedin.com/in/its-pulkit) | github.com/PulkitMangal09

PROFESSIONAL SUMMARY

Data Science undergraduate at IIT Madras with hands-on experience building, evaluating, and deploying machine learning, deep learning, and LLM-based applications. Skilled in Python, PyTorch, scikit-learn, Hugging Face Transformers, FastAPI, and Flask, with strong foundations in statistics, optimization, and algorithmic problem-solving. Experienced in developing end-to-end ML and NLP solutions, building Retrieval-Augmented Generation (RAG) pipelines and LLM-powered agents, and translating complex technical concepts into educational content for large audiences. Adept at combining analytical rigor with practical engineering to ship data-driven, AI-powered products at scale.

EDUCATION

Indian Institute of Technology, Madras
Bachelor of Science in Data Science and Applications

Chennai, Tamil Nadu
Sept. 2022 – Sept. 2026 (Expected)

EXPERIENCE

Data Scientist and SME, Scaler

Dec 2025 – Present

- Author scripts and lesson plans for the AI Engineering coursework, translating advanced topics (LLMs, RAG, and AI agents) into structured, accessible learning material.
- Prototype Retrieval-Augmented Generation (RAG) pipelines and LLM-powered agents to automate and streamline internal team workflows.
- Bridge the research-to-classroom gap, converting cutting-edge AI research into production-relevant curriculum delivered to hundreds of learners.

Technology Intern, Uncia (IITM Research Park)

Aug 2025 – Dec 2025

- Building machine learning and deep learning models for anomaly detection, association rule mining, and time series forecasting.
- Working on deployment and real-time streaming of models for production use cases.
- Contributing to the end-to-end ML pipeline from model development to operationalization.

Teaching Assistant, IIT Madras BS Degree Program

May 2023 – Present

Courses: Intro to Deep Learning & GenAI, Machine Learning Practice, Machine Learning Techniques, Mathematics-II

- Conducted interactive doubt-clearing sessions, supported students via course forums, and assisted in designing and evaluating assignments and quizzes.
- Simplified complex mathematical, deep learning, and generative AI concepts into intuitive explanations, enhancing student understanding across varied backgrounds.
- Designed coding exercises and guided participants through the mathematical intuition, vectorized implementation, and practical use-cases of core ML and DL models.
- Provided exposure to foundational GenAI concepts, linking modern applications with underlying machine learning principles.

PROJECTS

AI PSA Chatbot – Student Support Platform (Scaler)

2026

Production Project

- Built the full backend for an AI-powered student support chatbot serving Scaler learners across multiple course tracks (AIML, DSML, Full Stack Dev, DevOps, DSA).
- Designed and implemented a RAG pipeline using OpenAI text-embedding-3-small (1536-dim) with pgvector on PostgreSQL for knowledge base retrieval, and gpt-4o-mini for chat completion and intent classification.
- Engineered real-time streaming chat over Socket.IO with Redis for cross-instance fan-out, supporting concurrent conversations with token-level response streaming.
- Built an intent classification layer for persona routing (PSA, RM, Career) based on learner enrollment data, with configurable confidence thresholds for automatic human escalation.
- Implemented operator handoff system with role-based assignment, supervisor override, reassignment, and conversation resolution tracking (bot vs. human vs. awaiting).
- Developed admin APIs for knowledge base ingestion, blocked-topic management, and runtime-tunable app settings (thresholds, rate limits, handoff templates) without redeployment.
- Stack: FastAPI, PostgreSQL, pgvector, Redis, Socket.IO, OpenAI API, Alembic, Docker, Pytest.

DevOps & AI/ML Job Intelligence Agent

2025

Independent Project

- Built an end-to-end job intelligence platform scraping DevOps and AI/ML listings from 6 sources (Naukri, Indeed, LinkedIn, JSearch, Foundit, Instahyre) using Selenium, SeleniumBase, and REST APIs.
- Engineered a modular 9-stage cleaning and enrichment pipeline for deduplication, INR salary parsing, skill extraction, and experience-level classification.
- Implemented ML-based semantic curriculum alignment, scoring jobs against 17 AI Engineering modules using sentence-transformer embeddings (BAAI/bge-small-en-v1.5) and cosine similarity.
- Deployed a FastAPI REST backend with a PostgreSQL (Supabase) data store, a single-page frontend, and an automated scheduled pipeline.
- Github link: https://github.com/PulkitMangal09/job_search_agent

Automatic Disfluency Restoration (Kaggle)

2025

NPPE-2 Challenge

- Fine-tuned OpenAI Whisper (whisper-small) on a multi-modal (audio + text) task to reconstruct verbatim, disfluent transcripts from clean text and corresponding audio.
- Implemented prompt-conditioned ASR, steering the Whisper decoder with the clean transcript as a prompt while audio supplied acoustic evidence for hesitations and false starts.
- Optimized for Word Error Rate (WER) under Kaggle compute constraints; built modular training and inference pipelines using PyTorch and Hugging Face Transformers.
- Github link: <https://github.com/PulkitMangal09/Automatic-Disfluency-Restoration>

Growth Genie – Life Skills Platform

May 2025

Software Engineering Project

- Developed a full-stack life-skills web application for school children with a FastAPI backend and a Vue.js frontend, supporting student, parent, and teacher roles.
- Built modules for task management, financial-literacy tracking, emotional well-being logging, and health/nutrition monitoring.
- Integrated AI-powered features (dietitian chatbot and automated nutrition report generation) using an LLM API, with Supabase for database and authentication.
- Applied Pydantic for data validation and Pytest for backend testing as part of a collaborative team workflow.
- Github link: <https://github.com/PulkitMangal09/soft-engg-project-may-2025-se-May-12>

Ancient Texts Provenance Challenge (Kaggle)

October 2025

Deep Learning Practice Course Examination Project

- Developed an NLP pipeline to classify the geographical origin of ancient inscriptions, optimizing for the Macro F1 Score.
- Established a baseline model using TF-IDF vectorization and a Linear Support Vector Machine (SVM).
- Implemented and fine-tuned several Transformer-based models, including XLM-RoBERTa, DistilRoBERTa, and Ancient-Greek-BERT.
- Experimented with multiple fine-tuning techniques, including full model tuning, classifier-only tuning, and Parameter-Efficient Fine-Tuning (PEFT) using LoRA.
- Github link: https://github.com/PulkitMangal09/The_Ancient_Texts_Provenance_Challenge

Bank Telemarketing Success Prediction (Kaggle)

Oct 2024 – Nov 2024

Machine Learning Practice Course Project

- Built a supervised learning pipeline to predict customer conversion on a bank marketing dataset, with full preprocessing, EDA, feature engineering, and evaluation in scikit-learn.
- Compared multiple classifiers (Logistic Regression, SVM, XGBoost) to optimize for accuracy and F1 score.
- Github link: https://github.com/PulkitMangal09/mlp_project

ANN from Scratch

Jul 2025 – Present

Independent Project

- Building an artificial neural network from scratch in Python and NumPy, implementing activation functions, weight initialization, loss functions, and gradient descent to deepen understanding of forward/backward propagation.
- Testing on benchmark datasets (e.g., MNIST) to evaluate performance and improve generalization.
- Github link: https://github.com/PulkitMangal09/ann_from_scratch

Smart Factory Energy Prediction Challenge

March 2025 – April 2025

Independent Project

- Developed a regression model (Multi-layer Perceptron) to predict equipment energy consumption in a smart manufacturing facility from multivariate sensor and environmental time-series data.
- Identified key drivers of energy usage across factory zones and evaluated performance using MAE and RMSE, delivering actionable cost-reduction recommendations.
- Github link: https://github.com/PulkitMangal09/MLP_energy_prediction

TECHNICAL SKILLS

Languages: Python, Java, C, SQL (PostgreSQL, pgvector), JavaScript, HTML/CSS

ML & AI: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, XGBoost, LightGBM, sentence-transformers, OpenAI Whisper, LoRA/PEFT

AI Engineering: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), LLM Agents, Prompt Engineering, LLM Observability, NLP, Computer Vision, ASR, MLOps

MLOps & Cloud: MLflow, DVC, Docker, Kubernetes, Google Cloud Platform (GCP), CI/CD Pipelines, Model Serving

Data & Analytics: Pandas, NumPy, Matplotlib, Seaborn, Statsmodels, EDA, Feature Engineering, Time Series Forecasting, Anomaly Detection

Frameworks & Web: FastAPI, Flask, Streamlit, Vue.js, Socket.IO, Jinja2, SQLAlchemy, Pydantic, Celery, Redis, Requests, BeautifulSoup, Selenium

Developer Tools: Git, GitHub, VS Code, Jupyter Notebook, Google Colab, Postman, Vercel, Supabase, Alembic, Pytest