

Sandipan Paul

+91-7892710285 | psandipan20@gmail.com | [LinkedIn](#) | [GitHub](#) | [My Portfolio](#) | [LeetCode](#) | [Kaggle](#)

EDUCATION

Dr. Ambedkar Institute Of Technology

Bachelor of Engineering - Electrical & Electronics; CGPA: 8.13

Bengaluru, India

Dec. 2022 - Present

EXPERIENCE

AI/ML Intern

Strydden Technologies LLP.

June 2025 – Present

Ahmedabad, India — Remote

- Implemented **RAG** triage pipeline for clinical symptom matching-increased relevant retrieval rate by **20%**
- Built adaptive session management with **MongoDB** reducing session-loss inactivity by **90%**
- Integrated **WATI Whatsapp webhook** to enable Whatsapp triage, reducing user response time by **25%**
- Fixed embedding & RAG pipeline bugs & eliminated repeat failure cases by **90%**

Collaborating on a research paper under Prof. Harini Vaikund

Department of EEE, Dr. AIT

Dec 2024 – Present

Bengaluru, India

- Currently co-authoring an AI-driven power-systems** research paper.
- Performing**, end-to-end statistical data analysis, **developing**, and **validating** deep-learning models.

Intern

Just Robotics

Jul 2024 – May 2025

Bengaluru, India

- Engineered a RC remote-controller** using websockets and serial communication.
- Implemented a lightweight Flask backend** and **optimized** the controller interface by **30%**

PROJECTS

Observe | Python, Gradio, DVC, MLFlow, Huggingface, Grafana

Oct 2025

- Developed **Observe**, an end-to-end **MLOps** pipeline integrating FastAPI, Gradio, **DVC**, **MLflow**, **Prometheus**, and **Grafana** for **real-time NLP** model monitoring.
- Fine-tuned **DistilBERT** for news classification with custom tracking of inference latency, request metrics
- Designed and **containerized** full workflow for local-first observability, integrating experiment tracking, metric dashboards, and model versioning into a unified **Gradio UI** and implemented **Prometheus-based** telemetry collection and visualized latency, throughput, and confidence trends via **Grafana** dashboards.

Law-Vector | Python, Flask, Pinecone, BART, HuggingFace, Supabase, Streamlit

Jun 2025

- Developed an AI-driven legal** document search and understanding tool during Perplexity Hackathon, leveraging transformer-based **vector embeddings** and **Pinecone** for high-performance semantic retrieval.
- Implemented user **authentication** using Supabase, ensuring **90% uptime** and reducing unauthorized access incidents by **85%**.
- Designed and managed a modular **vector store** pipeline using sentence-transformers for embedding generation with **95% retrieval precision**.
- Applied a document **chunking strategy** to split large PDFs/texts into manageable sections, cutting average retrieval error rate by **30%** and improving relevance.
- Enabled **Retrieval-Augmented Generation** workflows by interfacing with the **Perplexity Sonar API**, boosting answer relevance scores by **25%**.

Paperdigest | Flask, Gemini API, Gensim, Sumy

Feb 2025

- Python-based research summarizer for papers and added **TTS** features for user accessibility.
- Utilized **Gensim** and **Gemini API** for summarization and implemented download and estimated features.

SKILLS

Languages: Python, C, C++, JavaScript, HTML/CSS

Tools & OS: AWS S3, AWS EC2, PostgreSQL, SQLite, SQL Workbench, MongoDB, Pinecone, Qdrant, ngrok, Git, Jupyter Notebook, Colab, GCP, Postman, Docker, MLFlow, Apache Airflow, Dagshub, DVC, Github Actions, Grafana

Libraries: Flask, FastAPI, Streamlit, Gradio, TensorFlow, Tesseract, Pytorch, OpenCV, NLTK, Spacy, Langchain

ACHIEVEMENTS

- Won the Innovator's Edge award for developing an interactive Braille script system (Hack Colossus 2024).