

# SUJIT SARKAR

AI Researcher | Multimodal AI · NLP · Applied ML

+91 9390573209 | [sujit.sarkar6112@gmail.com](mailto:sujit.sarkar6112@gmail.com) | [linkedin.com/in/sujit4070](https://www.linkedin.com/in/sujit4070) | <https://www.sujitsportfolio.com/>

## SUMMARY

Data Scientist and AI Software Engineer specialising in multimodal AI, NLP/OCR, and scalable ML pipelines. Experienced in shipping production AI systems, designing agentic workflows, and deploying cloud-native backends on GCP. IEEE-published researcher, international award recipient, and patent filer with hands-on experience leading cross-functional teams in research and product settings.

## CORE SKILLS

- NLP/OCR
- Machine Learning
- Prompt Engineering
- Agentic Workflow

## PROFESSIONAL EXPERIENCE

### AI/ML Engineer & Builder | Satyabodhi AI · Papergrader.in

July 2025 – Present

- Architected a multimodal AI grading pipeline, integrating 15+ research methodologies to maximize grading accuracy.
- Engineered a scalable, enterprise-grade multi-tenant backend on GCP Cloud Run, supporting 40–50 college tenants.
- Reduced evaluation error to RMSE 2.87 marks in the best pilot batch.

### Data Scientist (Contract - Granted Project)

Aug 2025 – Dec 2025

Bharath Research in AI and NextGen – KEM Hospitals

- Built an end-to-end ML pipeline, reducing 854 features to 30 through feature engineering and data analysis.
- Implemented and benchmarked 7+ ML models, achieving an RMSE of 218 g on the evaluation dataset.
- Presented findings at BrainCon 2026 (Singapore)

## TECHNICAL SKILLS

**Category1:** PyTorch · TensorFlow · scikit learn · Machine Learning · GPT · Gemini · LLaMA · Prompt Engineering · Generative AI · LLM Fine Tuning · PCA · AutoGen · Agentic Orchestration · Git · Jupyter · RAG

**Infrastructure & Backend:** Google Cloud Run (GCP) · FastAPI · Supabase · PostgreSQL · Redis · Docker

**Programming:** Python · JavaScript · Front End

**Databases:** PostgreSQL · Vector Databases (FAISS · ChromaDB)

**Analytics:** Power BI · Advanced Excel

## EDUCATION

### Master of Engineering, Data Science

Jul 2023 - Jun 2025

Army Institute of Technology, Pune

### Bachelor of Technology, Mechatronics Engineering

Jul 2019 - Jul 2023

SRM Institute of Science and Technology, Chennai

## KEY PROJECTS

### Dissertation Project: Intelligent Answer Sheet Evaluation

- Researched and validated 8+ reasoning and evaluation methods — spanning prompting strategies, self-refinement loops, and automated evaluation frameworks.
- Increased grading efficiency by ~30%; built a reasoning-driven automated pipeline and filed a patent.

### Adaptive AI Math Tutor - Multimodal RAG and Multi-Agent Learning System | Nov 2025

- Engineered a multimodal multi-agent Math Mentor using OCR, ASR, FAISS-based RAG, and human-in-the-loop validation.
- Built a FAISS feedback memory system to reuse solution patterns and user corrections, improving OCR/ASR accuracy.

### Optimizing Fake News Detection - Hybrid Linguistic + ML Approach | May 2024

- Processed Kaggle news dataset (42,677 articles) and engineered 32 linguistic features to enhance classification robustness.
- Evaluated XGBoost and SVM classifiers with PCA and cross-validation; published feature extraction code on GitHub.

## ADDITIONAL HIGHLIGHTS

- **IEEE PuneCon 2024** — Published research paper (NLP/ML domain)
- **IEEE PuneCon 2025** — Published research paper (AI evaluation systems)
- **BrainCon 2026, Singapore (MDIS)** — 1st Prize, Best Paper Award; 2 papers presented at international forum
- Indian patent filed for “System and Method for Multi-Agent AI–Based Evaluation of Handwritten Examinations” (PaperGrader).

## CERTIFICATIONS

- AI Agents - Agentic AI with Python Generative AI - Coursera, Apr 2025
- Prompt Engineering with Generative AI - Edureka, Dec 2024
- Advanced Excel & Power BI - Aparsoft Solutions, Apr 2024
- Front End Web Development - Cuvette, Dec 2023