

# SUDHANSU KAKKAR

Phone: 732-500-2147 | Email: [emsudhansu@gmail.com](mailto:emsudhansu@gmail.com) | LinkedIn: [linkedin.com/in/sudhansu-kakkar](https://linkedin.com/in/sudhansu-kakkar)

Portfolio: [emsudhansu.github.io/portfolio](https://emsudhansu.github.io/portfolio) | GitHub: [github.com/emsudhansu](https://github.com/emsudhansu)

## SUMMARY

Applied AI Engineer skilled in NLP, LLM fine-tuning, RAG, ML pipelines, with 5+ years of full-stack engineering experience.

## TECHNICAL PROFICIENCY

- AI / ML:** Python, PyTorch, TensorFlow, scikit-learn, Transformers, HuggingFace, BERT Fine-Tuning, NLP, Text Classification, LLMs, RAG Pipelines, Vector Databases (FAISS), LlamaIndex, LangChain, CrewAI, Prompt Engineering, Model Evaluation (F1, ROC-AUC, BLEU, ROUGE), Streamlit, Time-Series Forecasting (Prophet), Anomaly Detection
- Backend & Systems:** Java Spring Boot, Node.js, Flask, WebSockets (Socket.IO), SQLite, Microservices, REST APIs, PostgreSQL, MongoDB, AWS, CI/CD, Git, Apache Spark, ETL Pipelines,
- Frontend:** React, Redux Toolkit, HTML, CSS, JavaScript, Material UI, Micro-Frontends

## RELEVANT EXPERIENCE

### ▪ Research Assistant AI ML • STEVENS INSTITUTE OF TECHNOLOGY

May 2025 to Present

- Designed and fine-tuned transformer (BERT) to classify multi-level depression from text, achieving 81% accuracy.
- Built complete ML pipelines including dataset gathering, cleaning, preprocessing, tokenization, model training, validation, and hyperparameter tuning.
- Implemented evaluation workflows (F1-score, ROC-AUC, confusion analysis) and performed targeted error analysis to improve model stability and prediction consistency.

### ▪ AI Engineer and Developer • ONCREF

Jun 2025 to Jul 2025

- Built autonomous CrewAI agent workflows to automate web-scraping and summarization with minimal human supervision.
- Converted unstructured research webpages into structured JSON through LLM-based enrichment and classification steps to improve data usability and accuracy.
- Developed a lightweight end-to-end automation pipeline combining scraping, cleaning, and LLM-powered processing, reducing manual data preparation for the team.

### ▪ Senior SDE | Full Stack Developer • FIDELITY INFORMATION SERVICES (FIS)

Nov 2021 to Jan 2025

- Delivered major internet-banking features and standalone React + Spring Boot applications, supporting 400+ daily transactions and integrating with 15+ backend microservices across mission-critical systems.
- Improved engineering efficiency by introducing reusable UI components, micro-frontend patterns, and Node.js automation scripts, reducing manual build effort by 2+ hours/day and accelerating UI development cycles by ~50%.
- Enhanced system reliability and security by optimizing architecture, enforcing structured code reviews, and contributing to releases with reduced defect rates and strong 94% AppSec compliance.

### ▪ Front End Developer • TATA CONSULTANCY SERVICES (TCS)

Aug 2019 to Nov 2021

- Developed a cross-platform logistics app using React Native, Redux, and Material UI in collaboration with product teams.
- Ensured performance and stability through testing, debugging, deployment, and defect resolution across platforms.
- Improved user experience by fixing 200+ bugs and presenting weekly progress demos to stakeholders.

## EDUCATION

- **Master's in Applied Artificial Intelligence** — Stevens Institute of Technology, United States — 2025 to 2026 (Expected)
- **Bachelor's in Computer Science Engineering (CGPA: 8.4)** — GGSIPU, India — 2015 to 2019

## PROJECTS

- **TruthfulQA Evaluation Pipeline — LLM Truthfulness Scoring** — (2025)
  - Evaluated GPT-4o-mini factual accuracy on TruthfulQA using DistilBERT truth classifier, achieving ~84% truthfulness.
- **Multi-Class Depression Detection using BERT** — (2025)
  - Transformer-based multi-class classifier predicting depression severity from social media text with ~81% accuracy.
- **Air Quality Index Forecasting & Anomaly Detection** — (2025)
  - Processed 29k+ AQI records with a Spark pipeline. Used Prophet to forecast PM2.5 with automated anomaly detection.
- **UniChat - Multilingual Chat with Sentiment AI** — (2025)
  - Chat platform with auto-translation, custom TF-IDF sentiment model, and SQLite history using Flask and WebSockets.
- **AI Resume Chatbot Twin (RAG + Llama 3)** — (2025)
  - Llama-3 based RAG chatbot using all-MiniLM embeddings and JSON resume data stored in a persisted vector index.
- **Dynamic Attendance Manager** — (2019)
  - Role-based attendance system using React, Redux-Saga, and Material UI for students, teachers, and admins.

## CERTIFICATIONS AND TRAININGS

- **Generative AI: Working with Large Language Models** — LinkedIn Learning (2025)
- **AI Coding Agents with GitHub Copilot & Cursor** — LinkedIn Learning (2025)