Boddu Sai Ganesh Reddy

9392359162 | boddusaiganesh81@gmail.com | github.com/boddusaiganesh | Andhra Pradesh, India

Professional Summary

AI/ML Engineer with hands-on experience developing production-ready intelligent systems using Generative AI, Large Language Models, and advanced machine learning techniques. Expertise in building multi-modal AI applications, Agentic AI systems with multi-agent orchestration, RAG pipelines, computer vision, and NLP systems. Strong foundation in deep learning frameworks (PyTorch, Hugging Face Transformers), natural language processing, and full-stack development. Recognized as GHCI 2025 Hackathon Finalist for innovative AI-powered verification system achieving 99% performance improvement through intelligent automation and fraud detection algorithms.

Education

VIT-AP University — Amaravati, India
Bachelor of Technology in Computer Science and Engineering, CGPA: 8.88/10 (88.8%) — *Sept 2022 – Apr 2026*

- Expected Graduation: June 2026
- Relevant Coursework: Machine Learning, Deep Learning, Artificial Intelligence, Natural Language Processing, Computer Vision, Data Structures and Algorithms, Statistics and Probability, Database Systems

State Board Senior Secondary Education — India
Class XII, Score: 94% — *2022*

Technical Skills

AI/ML Frameworks: PyTorch, Hugging Face Transformers, SpaCy, rembg (U-2-Net)
Generative AI: Large Language Models (LLMs), Google Gemini API, Groq API (Llama 3), Prompt Engineering, BERT, RAG Pipelines, Model Fallback Strategies
RAG & Vector Databases: Retrieval Augmented Generation (RAG), ChromaDB, FAISS, Semantic Search, Document Chunking, Sentence-Transformers Embeddings
Agentic AI: Multi-Agent Systems, Agent Orchestration, Chain-of-Thought Reasoning, Self-Critique Agents, Risk Detection Agents
AI Specializations: Multi-Modal AI, AI Background Removal, Rule-Based Compliance Engines, Guardrails & Evaluation Systems
Natural Language Processing: Text Classification, Named Entity Recognition (NER), Dependency Parsing, Coreference Resolution, Part-of-Speech Tagging, Document Understanding, Information Extraction, AI Text Compliance Validation
Computer Vision: OCR (Optical Character Recognition), Facial Recognition, Liveness Detection, AI Background Removal, Image Processing with Pillow
ML Techniques: Supervised Learning, Feature Engineering, Custom Transformer Models, Hybrid AI Architectures
Programming Languages: Python (Advanced), JavaScript, TypeScript, Java, SQL, HTML5, CSS3
Python ML/Data Libraries: NumPy, Pandas, NLTK, Coreferee, Pillow, OpenCV, Pydantic, ChromaDB, Sentence-Transformers
MLOps and Deployment: Docker, FastAPI, Streamlit, Uvicorn
Backend Development: Python FastAPI, Node.js, Express.js, Java Spring Boot, RESTful API Design, JWT Authentication, Rate Limiting, CORS
Frontend Development: React, TypeScript, Tkinter, Vite
Databases: PostgreSQL, MySQL, SQLite, SQL Optimization, ChromaDB (Vector Store)
Cloud and APIs: Google Cloud AI, Docker containerization, Twilio, SendGrid
Development Tools: Git, GitHub, VS Code, Postman, Maven

Professional Experience

Java Developer Intern — Aug 2025 – Oct 2025
Infosys Springboard — *Remote*

- Engineered and launched MedVault, a full-stack hospital management web application with AI-powered features serving multiple user roles including administrators, doctors, and patients.
- Developed secure stateless REST API using Java Spring Boot with JWT-based authentication and role-based access control implementing machine learning-ready architecture for future predictive analytics integration.
- Built dynamic responsive user interface with ReactJS creating multi-page dashboard for appointment booking, patient management, and feedback collection with potential for ML-driven recommendations.
- Designed and optimized MySQL database schema using JPA and Hibernate to manage complex entity relationships preparing data pipelines for machine learning model training on patient and appointment data.

- Implemented comprehensive data preprocessing and feature engineering workflows handling medical records enabling future integration of NLP models for automated patient data extraction.
- Collaborated with cross-functional teams following agile methodology while exploring opportunities to integrate AI/ML solutions including predictive scheduling and intelligent resource allocation.

Projects

AI-Powered KYC Verification System with Real-Time Fairness Monitoring 2025

- *GHCI 2025 Hackathon Finalist*
- **Technologies:** Node.js, Express, React, TypeScript, PostgreSQL, Google Gemini AI, Twilio, SendGrid, JWT, Docker
- Architected production-ready KYC platform reducing identity verification time from 2-7 days to 5-10 seconds with 99% faster processing and 95% cost reduction per verification.
- Engineered industry-first bias detection system with live EEOC 80% Rule fairness monitoring and real-time bias alerts ensuring equitable approvals across demographic groups.
- Integrated Google Gemini AI for multi-modal document OCR, facial recognition, liveness detection, and parallel fraud assessment across 6 verification dimensions.
- Built 8-step automated KYC workflow with contact validation, OTP verification, document processing, address proof, risk assessment, and fraud detection.
- Implemented comprehensive admin dashboard with real-time analytics, bias/fraud monitoring, full audit trails, and compliance-ready data exports for regulatory adherence.
- Designed scalable PostgreSQL database with 12+ strategic indexes and materialized views supporting 10,000+ daily applications with optimized query performance.
- Developed 6-layer security architecture including TLS/HTTPS, JWT authentication, AES-256 encryption, CORS protection, and complete audit logging for GDPR/KYC/AML compliance.

AI-Powered MD&A Generator - Automated Financial Report Generation 2026

- **Technologies:** Python, Google Gemini API, Groq (Llama 3), ChromaDB, RAG, Pandas, FastAPI, Streamlit, Pydantic, HuggingFace Embeddings
- Architected Agentic AI system with multi-agent orchestration (Critique Agent, Risk Detection Agent, Comparative Analysis Agent) for automated Management Discussion & Analysis report generation from SEC financial statements.
- Engineered RAG pipeline using ChromaDB vector store with local HuggingFace embeddings (sentence-transformers/all-MiniLM-L6-v2) achieving zero-cost retrieval and context-grounded LLM outputs.
- Built comprehensive KPI calculation engine computing 14+ financial metrics including YoY/QoQ growth rates, profit margins, liquidity ratios, and automated trend detection.
- Implemented 6-layer guardrails evaluation system validating factual consistency (5% tolerance), citation coverage, metric accuracy, financial reasonableness, and content quality.
- Designed rate-limit resilient architecture with automatic Gemini model rotation and Groq (Llama 3) API fallback ensuring uninterrupted generation during API quota exhaustion.
- Developed dual interfaces including Streamlit interactive dashboard for demo and FastAPI REST service with async job processing for production scalability.
- Achieved 95%+ factuality scores through RAG-grounded generation with automatic citation linking to source financial data chunks.

Tesco Creative Studio - AI-Powered Retail Media Creative Tool 2025

- *Tesco Retail Media Hackathon*
- **Technologies:** React 18, Vite, Fabric.js, Python FastAPI, Google Gemini AI, rembg (U-2-Net), Pillow, Zustand, SQLite
- Developed full-stack AI-powered creative builder enabling advertisers to autonomously create guideline-compliant retail media creatives with real-time compliance validation.
- Engineered AI background removal service using rembg (U-2-Net deep learning model) for one-click product packshot extraction with high-quality results in 2-5 seconds.
- Built comprehensive 18-rule compliance validation engine implementing Tesco brand guidelines covering copy rules, design specifications, accessibility (WCAG AA contrast), and alcohol advertising requirements.
- Integrated Google Gemini AI for intelligent text compliance checking, AI-powered layout suggestions generating 5 layout variations, and channel-optimized campaign set generation.
- Developed interactive canvas editor using Fabric.js with drag-and-drop interface, multi-layer element management, and real-time rendering supporting 4 export formats (1:1, 9:16, 1.91:1, 4:5).
- Implemented multi-format export service with automatic image optimization to under 500KB, supporting JPEG/PNG outputs with intelligent quality compression.

- Built Zustand-based state management system handling canvas state, element tracking, compliance results, and real-time validation feedback with 16+ React components.
- Designed RESTful API with FastAPI featuring rate limiting (20 req/min), CORS middleware, and comprehensive endpoints for file upload, AI services, compliance validation, and export functionality.

Crime Information Extraction and Intensity Identification System 2025

- **Technologies:** Python, PyTorch, BERT, Hugging Face Transformers, SpaCy, Coreferee, Streamlit, Tkinter, NLP
- Developed unified crime text analysis platform combining deep learning intensity prediction with rule-based entity extraction achieving comprehensive crime report understanding.
- Engineered custom BertWithFeatures model integrating BERT transformer embeddings with 6 handcrafted SpaCy features (trigger score, modifiers, weapons, locations, dates) for crime severity classification on 1-5 scale.
- Built sophisticated NLP pipeline using SpaCy en_core_web_lg for Named Entity Recognition, dependency parsing, and part-of-speech tagging to extract perpetrators, victims, locations, dates, times, and weapons.
- Implemented coreference resolution using Coreferee library for pronoun disambiguation and entity linking, enabling accurate role assignment across multi-sentence crime narratives.
- Developed intelligent RoleAnalyzer module using hierarchical rule-based system analyzing grammatical dependencies (nsubj, nsubjpass, dobj) to distinguish perpetrators from victims with high precision.
- Created EventExtractor component reconstructing Agent-Action-Patient event structures from crime text supporting both active and passive voice patterns with chained verb handling.
- Built dual-interface application with Streamlit web UI and Tkinter desktop GUI featuring real-time analysis, preprocessing visualization, and academic evaluation metrics (Precision, Recall, F1-Score).
- Designed modular architecture with 15+ specialized modules including EntityRefiner for post-processing deduplication and ResolutionEngine for canonical entity resolution.

Achievements and Recognition

GHCI 2025 Hackathon Finalist

- Selected among top finalists for AI-Powered KYC Verification System demonstrating exceptional AI/ML problem-solving skills, innovative approach to bias detection, and production-ready system design.

Academic Excellence

- Maintained 8.88/10 CGPA (88.8%) with strong performance in Computer Science and Engineering coursework demonstrating solid foundation in algorithms, data structures, and software development.

Gold Medal - Inter-University Martial Arts Championship (2024)

- Demonstrated discipline, perseverance, dedication, and ability to perform under competitive pressure.

Bronze Medal - State-Level Karate Tournament (2023)

- Showcased commitment to excellence and resilience in competitive environment.

Leadership and Activities

Club Manager, Admin Team Jun 2024 – Dec 2025
Dragon Fist Martial Arts Club, VIT-AP University *Amaravati, India*

- Managed diverse team of 150+ members and coordinated dynamic events ensuring smooth club operations and high member engagement.
- Led strategic planning sessions and delegated tasks effectively to drive organizational goals and increase member participation by 40%.
- Demonstrated strong communication and organizational skills coordinating with university administration and external stakeholders.

Team Member 2023 – Present
University Karate Team, VIT-AP University

- Secured gold and bronze medals in multiple inter-university martial arts tournaments demonstrating discipline and competitive excellence.

Certifications

MERN Stack Development Certification Course
Java Foundation Certification
Introduction to Web Development with HTML5, CSS3, and JavaScript