

Sai Sri Swetha Battu

New York , US | sabattu@my.bridgeport.edu | 740-602-7120 | yourwebsite.com

linkedin.com/in/saisri-swetha-battu-889a0649 | github.com/Saisriswetha

Professional Summary

GenAI Engineer with 3+ years of experience and specialized in generative models and natural language processing, with hands-on expertise in LLMs (GPT, LLaMA, Falcon), MLOps, cloud deployment (AWS, Azure OpenAI), and data pipeline development. Demonstrated expertise in building Model Context Protocol (MCP) and Agent-to-Agent (A2A) integrations, as well as CrewAI frameworks for advanced automation. Adept at driving end-to-end AI initiatives, from research and prompt engineering to model optimization, project management, and stakeholder communication. Passionate about ethical AI and delivering robust, production-ready solutions that support business goals. Strong technical foundation in Python, JavaScript, SQL, and cloud platforms (AWS, GCP, Azure), alongside containerization (Docker, Kubernetes) and CI/CD best practices. Proven track record of boosting system efficiency by 25%, optimizing real-time data processing, and establishing end-to-end model observability.

Projects/Applied Experience

Cognitive HOI Reasoning – Research Project

Apr 2023 – Mar 2024

- Investigated human-object interaction models on **Bongard-HOI** and **HAKE datasets**, identifying reasoning gaps between synthetic and real-world image understanding.
- Applied **DVRL** and **Componential Analysis**, raising model accuracy from 55% to 80%.
- Contributed novel insights into bridging cognitive paradigms in computer vision.

Personal Attorney – Legal RAG Assistant

Apr 2024 – Nov 2024

- Built a **Retrieval-Augmented Generation (RAG)** system to provide state-specific DMV/legal information, reducing expensive LLM retraining costs.
- Implemented with **ChromaDB**, **LangChain**, **FastAPI**, **DeepSeek 8B (LoRA fine-tuning)**, enabling scalable updates without retraining.
- Delivered practical support for immigrants navigating complex legal frameworks.

SmartMed AI – Pill Detection App

Jan 2025 – Aug 2025

- Developed a computer vision pipeline using **YOLO (v5–v12)**, multi-scale detection, attention modules, and frame-to-frame tracking to handle overlapping pills.
- Deployed solution for pharmacies, improving operational accuracy and reliability in pill recognition tasks.
- Tech Stack: **PyTorch**, **OpenCV**, **ResNet**, **FastAPI**, **AWS**.

Time-Series Forecasting – Stock Market Analysis

Sep 2025 – Running

- Designed forecasting models for volatile financial time-series data.
- Compared traditional **ARIMA/SARIMA** with advanced methods like **Meta’s Prophet** and **XGBoost ensembles**.
- Improved accuracy in predicting market fluctuations under high volatility conditions.

Technologies

Programming Languages: Python | Java | C/C++ | JavaScript | SQL | NoSQL | Algorithms & Data Structures

Machine Learning & AI: Model Training | Fine-Tuning | Deployment | Optimization | Large Language Models (LLMs) | Generative AI (GenAI) | Deep Learning | NLP | Computer Vision | Prompt Engineering | GPT | LLaMA | Falcon | Deep Learning | Neural Networks | Machine Learning | NLP | Text Generation | Sentiment Analysis | Named Entity Recognition | Retrieval-Augmented Generation | Synthetic Data Generation | Model Fine-Tuning | Model Optimization | Bias Control | Ethical AI

Frameworks & Cloud Platforms: PyTorch | TensorFlow | Keras | scikit-learn | Hugging Face Transformers | LangChain | FastAPI | Flask | Pandas | NumPy | SQLAlchemy | AWS (S3, EC2, SageMaker, Bedrock) | Azure ML, OpenAI | GCP AI/ML | Docker | Kubernetes | CI/CD | MLflow | Kubeflow | MLOps | Cloud Deployment | Microservices

Strengths & Tools: Git | GitHub | Postman | PyCharm IDE | ITSM Tools | Problem Solving | Debugging | Clean Code | Object-Oriented Programming | Design Patterns | Collaboration | Cross-functional Collaboration | Attention to Detail

Education

University of Bridgeport, M.S. in Computer Science Aug 2023 – Dec 2025

- Focus on Artificial Intelligence, Full-Stack Development, and Cloud Computing
- Relevant Coursework: Advanced Machine Learning, Cloud Computing, Software Engineering

Adikavi Nannaya University (AKNU), BCom with Computer Science Jun 2019 – Jun 2023

- Completed coursework in Data Structures, Algorithms, Operating Systems, and Databases
- Academic Projects: Open Banking API Aggregator for Seamless Financial Data Integration

Experience

Starrez Developer, University of Bridgeport – Bridgeport, CT Aug 2024 – Apr 2025

- Managed and customized the StarRez housing portal to support student housing operations across semesters.
- Developed and deployed StarRez Wizards to streamline housing applications, automate deposit processing, and support semester-specific procedures, reducing manual workload and errors.
- Integrated housing workflows with AwareManager and UpKeep portals, accelerating maintenance coordination and facilities management response times.
- Built automated tools for dynamic reporting and portal updates using Python and AWS, which boosted operational efficiency and enabled real-time insights.
- Collaborated with cross-functional campus departments to optimize data integration and elevate user experience across multiple platforms.