KEERTHANA VANGALA

vangala.keerthana20@gmail.com | +1 929 662 1511 | | [LinkedIn](#) | [Portfolio](#)

Education

State University of New York at Buffalo, USA — *August 2024 – December 2025*
M.S. in Computer Science — **GPA: 3.54/4.00**
Courses: *Machine Learning, Data-Intensive Computing, Statistical Data Mining, Database Systems (Capstone), Algorithms Analysis*

G. Narayanamma Institute of Technology and Science, India — *July 2018 – June 2022*
B.Tech. in Electronics and Communication Engineering — **GPA: 8.5/10.0**
Courses: *Python, Cloud Computing, Computer Security, Data Structures and Algorithms, Operating Systems*

Skills

Languages:	Python, R, C, C++, SQL, Java, Bash, React, PySpark, Mainframe (COBOL), HTML, CSS
Data Engineering:	AI/ML, PyTorch, Apache Spark, Hadoop, Kafka, Hive, Airflow, Databricks, ETL Pipelines
Cloud & Infrastructure:	AWS (EC2, S3, Lambda, QuickSight, Redshift, SageMaker, Amplify), Docker, CI/CD, GitHub
Databases & Storage:	MySQL, MongoDB, Redis, Data Warehousing (Redshift, Snowflake), Supabase
Data Visualization	Tableau, Power BI, QuickSight, Plotly, Streamlit

Experience

Resilience — *Remote, US*
Software Development Intern — *May 2025 – August 2025*
- Designed and deployed an AI-enabled **Hazard & Security Risk Assessment (HVA/SRA)** web app using **React**, enabled real-time risk evaluation across **t**eams.
- Integrated **Supabase (Auth, Postgres)** with dynamic relational models for secure storage and **real-time analytics**, boosting data retrieval speed by **40%**.
- Implemented RESTful APIs in Python and automated CI/CD pipelines via AWS Amplify, reducing deployment time by 30% in collaboration with a 4-member team.**Tech Stack:** React, Supabase, Python, AWS Amplify, CI/CD, REST APIs

Deloitte — *Hyderabad, India*
Analyst — *September 2022 – August 2024*
Client-Vanguard:
- Migrated legacy **COBOL+DB2** mainframe applications to the **AWS Cloud** using **PySpark ETL**, reducing infrastructure cost **25%**.
- Automated dependency mapping for hundreds of mainframe jobs via metadata parsing, cutting manual tracing **85%**.
- Developed scalable **AWS Glue** ETL workflows integrating DB2 outputs with **S3, Redshift, QuickSight**, processing **500+ jobs/day**.**Tech Stack:** Python, PySpark, AWS Glue, DB2, COBOL, S3, Redshift, QuickSight, Step Functions, ETL Pipelines
Client-Elevance Health:
- Engineered the **"Timber"** AI-driven claim-verification platform automating validation of **ICD-10** and **CPT/HCPCS** codes.
- Integrated **COBOL + DB2** backend with an AI-based rule engine detecting **2000+** billing and claim inconsistencies monthly. and tested **edit-trigger logic** for batch claim pricing and compliance across **500+ batches** within payer networks.
- Created automated **Postman API** collections to simulate batch claims and ensure end-to-end data consistency.
- Optimized workflows, cutting review time **30%** and saving **$16.8K/month**; visualized analytics via **QuickSight, React, Power BI**.
- **Tech Stack:** Python, COBOL, DB2, React, AWS QuickSight, Power BI, Postman, ICD-10/CPT/HCPCS Validation, AI Rule Engine, Claim Edit Automation

Projects

Noti Research Mini – Agentic Multi-Agent Research Assistant [GitHub](#) — *August 2025 – October 2025*
- Designed an **agentic AI system** integrating planner, summarizer, and citation agents via **LangChain** for research automation.
- Architected a modular **FastAPI backend** with JWT-based authentication, **SQLAlchemy+ SQLite** persistence, and orchestrated agent execution.
- Developed a lightweight **HTML/JS** front end for interactive research querying and visualization.
- **Tech Stack:** Python, FastAPI, SQLAlchemy, SQLite, JWT, bcrypt, LangChain, OpenAI API, Docker

Maternal Infant Insight – LLM-Augmented Risk Prediction — *October 2024 – January 2025*
- Programmed an **end-to-end ML pipeline** using **scikit-learn, XGBoost, pandas** with integrated **EDA**, **feature correlation analysis**, and statistical validation for maternal health data.
- Benchmarked **Logistic Regression, Random Forest, XGBoost** models (F1 = **0.94**, Recall = **0.93**); applied **cross-validation** and **SHAP** for interpretability and feature importance ranking.
- Integrated an **LLM-based explainer** to automatically summarize key risk drivers for clinicians.
- **Tech Stack:** Python, scikit-learn, XGBoost, SHAP, Large Language Models, NumPy, Matplotlib, EDA

Predictive Loan Advisor – Machine Learning Capstone — *October 2022 – November 2022*
- Orchestrated an **ETL pipeline on Databricks (PySpark)** processing **50K+** loan records, performing feature encoding, scaling, and correlation analysis to streamline model-ready data for ML training.
- Identified key drivers: **Debt-to-Income Ratio, Credit History, LTV**, improving prediction accuracy.
- Trained and tuned **XGBoost, Random Forest, Logistic Regression** on **AWS SageMaker** (**AUC = 0.96**).
- Built dashboards in **AWS QuickSight** and **Power BI** showing default trends and risk metrics.
- **Tech Stack:** Databricks, PySpark, SageMaker, S3, Aurora, XGBoost, QuickSight, Power BI, Python

Certifications

- **AWS Certified Cloud Practitioner (CLF-C01)** & **AWS Certified Developer – Associate (DVA-C02)**
- Completed **Postgraduate Program in Software Engineering and Data Science** from Great Learning with **IIIT Hyderabad**