

Sai Amulya Pingili

Tempe, Arizona | +16027919546 | saiamulyapingili@gmail.com | <https://www.linkedin.com/in/amulyapingili/>

EDUCATION

Arizona State University

Master of Science, Data Science, Analytics and Engineering

Tempe, Arizona

Expected May 2026

TECHNICAL SKILLS

Languages: Python, R, C++, Java

Databases: PostgreSQL, MongoDB, MySQL, Amazon Redshift, Microsoft SQLServer, Amazon DynamoDB

Tools & Platforms: Apache Airflow, Pandas, Matplotlib, ChromaDB, SentenceTransformer, Hugging Face, Git

ML/NLP Tools: Hugging Face, SentenceTransformers, Topic Modeling

BI Tools: Power BI, Tableau

Cloud: AWS(S3,IAM,EC2)

EXPERIENCE

ASU Enterprise Technology

AI Data Science Intern

Scottsdale, Arizona

June 2025 – Present

- Evaluated ethical benchmarks across 40+ large language models (OpenAI, Anthropic, Amazon, Meta) and built a repeatable workflow to track results across model versions.
- Developed an LLM-based topic modeling workflow for the ASU Experience Centre to surface recurring student concerns during peak periods and supported a chatbot for answering common questions.
- Built an automated Python pipeline to generate evaluation reports and store outputs, reducing manual reporting time by 60%.
- Annotated 200+ Mastercard Foundation internal documents with structured metadata to support downstream fine-tuning and knowledge base development.
- Designed an initial natural-language-to-SQL capability for self-service analytics to help stakeholders query platform data on users, projects, and usage.

Excelerate

Data Visualization Intern

Remote, India

July 2023 – Aug 2023

- Analyzed performance of 10 Facebook ad campaigns using Python and Tableau, evaluating 8+ KPIs including CPC, CPR, reach, impressions, and engagement metrics.
- Identified 20% underperforming campaigns by diagnosing high cost and low engagement, uncovering major budget inefficiencies.
- Recommended campaign elimination and budget reallocation strategies to improve overall ROI through data-driven decision-making.

PROJECTS

NLQ Traffic Analytics Dashboard

Sep 2025 – Dec 2025

- Designed and implemented a data engineering pipeline using Apache Airflow to ingest NYC Open Data, perform data cleaning, geohash computation, and batch loading into a distributed MongoDB cluster.
- Validated ETL logic across 2M+ crash records, ensuring data quality, shard balance, and spatial-temporal query consistency under concurrent workloads.
- Integrated embedding generation during ETL to enable semantic and natural-language query support using vector search.

RAG Movie Recommender Chatbot

March 2025 – May 2025

- Built a low-latency Retrieval-Augmented Generation pipeline using ChromaDB and SentenceTransformer embeddings (6k+ movies), optimizing vector search paths to return semantic matches in under 150ms.
- Integrated Google's Gemini API for contextual response generation, improving recommendation relevance by 28% through prompt-conditioning and retrieval-aware generation.
- Developed a real-time Gradio interface with async request handling and caching layers, enabling smooth multi-turn conversations and reducing model inference overhead during peak usage.

Predicting Alien Worlds (Live Demo)

Aug 2023 – June 2024

- Improved exoplanet habitability prediction accuracy by integrating NASA's Kopparapu HZ equations with Scikit-Learn models, reducing false-positive classifications across 6,000+ confirmed exoplanets.
- Built an interactive planetary simulator enabling real-time tuning of orbital and stellar parameters, increasing model interpretability and user engagement through dynamic habitability recalculations.
- Optimized full-stack performance by modularizing Python logic and implementing native static asset serving, cutting page load times and ensuring smooth deployment on Streamlit Cloud.

CERTIFICATIONS

DP-900: Azure Data Fundamentals (Microsoft)
Problem Solving Basics — HackerRank