ANKITA SETHI

Sunnyvale, CA | +1(530) 407-7394 | ankitasethi282@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL EXPERIENCE

Graduate Research Assistant| *Stony Brook University| Stony Brook, NY*　　　　　　　　　　July 2025 – Present
- Engineered an end to end content segmentation system using **Python**, **Django** and **Redis** caching, reducing response latency by 35% and improving segmentation stability across diverse pages.
- Designed asynchronous data processing pipelines and analyzed throughput and error patterns to uncover optimization opportunities.
- Integrated **Gemini 2.5** and **LangChain** into the segmentation workflow and ran controlled evaluations that improved accuracy by 18%, using prompt versioning, experiment tracking & model comparison to enable faster iteration and data backed model selection.

Graduate Research Assistant| *Stony Brook University| Stony Brook, NY*　　　　　　　　　　Aug 2024 – May 2025
- Led ML driven accessibility research by integrating **LLMs** with NVDA to interpret documents, charts and web content, analyzing patterns in model errors and user corrections to improve comprehension.
- Developed Chrome extensions and **NVDA** plugins connected to Django pipelines and used structured metrics such as latency trends, failure categories and accuracy drift to evaluate performance.

Data Engineer | *Accenture Private Limited | Bangalore, India*　　　　　　　　　　　　　　Jun 2021 – Aug 2023
- Engineered large scale **ETL** pipelines using **Talend** and **Snowflake**, increasing data throughput by 35% and reducing end to end pipeline latency across production.
- Streamlined **BERT** based **NLP** classifiers for pharma workflows, improving entity recognition accuracy and automating structured labeling across more than 2 million unstructured text records.
- Built 15+ **Tableau** dashboards for pipeline health, anomaly spikes & data quality trends, cutting issue identification time by 40%.
- Created a **Flask** chatbot that handled 50+ daily service tickets, reducing misclassification and improving user support workflows.
- Optimized **SQL** queries and Snowflake schemas with validations, alerts, reducing compute cost & removing manual QA.
- Incorporated **CI/CD** pipelines using Jenkins, enabling automated testing, reliable rollbacks and consistent weekly prod deployments.
- Collaborated with global teams, mentored 10+ offshore engineers, improving delivery consistency through **Agile** and **Jira** workflow.

TECHNICAL SKILLS

Machine Learning & Statistical Methods: Feature Engineering, Classification, Regression Analysis, Time-Series Forecasting (ARIMA), Anomaly Detection, NLP, Deep Learning(CNN, Transformers), Clustering, Sampling Techniques, Hypothesis Testing & Confidence Intervals, Principal Component Analysis, Data Visualization, Statistical Inference, A/B Testing.
Programming & Tools: Python (NumPy, Pandas, Scikit-learn, Matplotlib, PyTorch, spaCy, NLTK, OpenCV, Streamlit), SQL, Snowflake, PostgreSQL, Java, C, C++, Git, Tableau, Talend,, AWS, GCP, HTML, CSS.
Data Engineering & Platforms: ETL Pipelines, Data Modeling, Large Scale Data Processing, Redis Caching, Prometheus, Grafana.
Deployment & DevOps: CI CD Pipelines, Docker, Jenkins, Linux, MLOps Concepts, Jira, ServiceNow, Agile.

PROJECTS AND RESEARCH

Member Aware Question Answering System　　　　　　　　　　　　　　　　　　　　　　　　Nov 2025
- Built a retrieval-augmented QA system using Gemini 2.5, Django and JSON retrieval for more than 3.5K member messages.
- Implementing context ranking, update detection and prompt construction with median latency under 2 seconds.
- Structured scalable pipeline for message extraction, feature preparation, deployed the service on Render with ongoing latency tracking.
- *Python, JSON, JavaScript, Render, Postman, LLM*

Multi-Agent Approach for Detecting Hallucination in LLMs　　　　　　　　　　　　　　　　Sep 2024
- Designed a multi-agent framework to detect hallucinations in LLM outputs using concept inference and NER, improving factual reliability by 18% in evaluation benchmarks.
- Scaled domain coverage from 7 to 57 via self-familiarity scoring, increasing detection consistency across varied text types by 22%.
- *Named Entity Recognition (NER), Concept Inference, Hugging Face Transformers*

Smart Energy Assistant for Smart Homes　　　　　　　　　　　　　　　　　　　　　　　　Feb 2024
- Engineered a home energy assistant using ARIMA model to forecast appliance usage based on time-series and weather trends, reducing prediction error by 28% across 1000+ observations.
- Integrated GPT-4 for anomaly detection and achieved $R^2 = 0.737$ through walk-forward validation.
- *Time-Series Forecasting (ARIMA), GPT-4, Feature Engineering, Anomaly Detection*

EDUCATION

Stony Brook University | Stony Brook, NY　　　　　　　　　　　　　　　　　　　　　　Aug 2023 - May 2025
Master of Science in Computer Science
Relevant Courses: *Data Science, Visualization, Recent progress in AI/ML, Analysis of Algorithms*

Gandhi Institute of Technology and Management - GITAM University | Visakhapatnam, India　　Jun 2017 - Jun 2021
Bachelor of Technology in Computer Science and Engineering