

# ANKITA SETHI

Sunnyvale, CA | [ankitasethi282@gmail.com](mailto:ankitasethi282@gmail.com) | [LinkedIn](#) | [GitHub](#)

## PROFESSIONAL EXPERIENCE

- Graduate Research** | *Stony Brook University* | *Stony Brook, NY* July 2025 – Present
- Advanced research on ML enhanced accessibility by building an end-to-end pipeline for web content segmentation to improve NVDA-based navigation and leveraging Gemini 2.5 with Langchain prompt templates for accurate segmentation.
  - Python, Django, NVDA, Gemini 2.5, LangChain.*
- Graduate Research** | *Stony Brook University* | *Stony Brook, NY* Aug 2024 – May 2025
- Conducted independent research on ML enhanced accessibility systems, integrating LLMs with NVDA to interpret documents, charts and web content, boosting comprehension for visually impaired users.
  - Delivered a year long advanced project and independent study on object identification, task automation, LLM experimentation, reusable data pipelines, Chrome extension development, NVDA plugin-based navigation, and Django-based server integration.
  - PyTorch, NVDA, Large Language Models(Gemini, ChatGPT APIs, LangChain), OpenCV, Python, Django.*
- Data Engineer** | *Accenture Private Limited* | *Bangalore, India* Jun 2021 – Aug 2023
- Architected scalable ETL pipelines using Talend and Snowflake, improving data throughput and reducing latency across production.
  - Optimized BERT-based NLP classifiers for pharma workflows, streamlining entity recognition and structured labeling across large-scale unstructured datasets.
  - Developed and deployed 15+ monitoring dashboards in Tableau and Power BI to track pipeline health, flag anomalies, and support post-release diagnostics.
  - Engineered a Flask chatbot that handled 50+ daily service tickets, reducing misclassification and improving user support workflows.
  - Rebuilt SQL and Unix pipelines with validations and alerts, boosting data accuracy and removing manual QA in staging and prod.
  - Integrated CI/CD pipelines via Jenkins, enabling automated testing, rollback support and consistent weekly production deployments.
  - Collaborated on 5+ cross-functional initiatives and mentored 10+ offshore engineers across 5+ time zones, accelerating onboarding by 50% and driving delivery using Agile and Jira workflows.

## 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.

**Programming & Tools:** Python (NumPy, Pandas, Scikit-learn, Matplotlib, PyTorch, spaCy, NLTK, OpenCV, Streamlit), SQL, Java, C++, Git, Talend, Snowflake, Tableau, MS Excel, Power BI, AWS.

**Deployment & DevOps:** CI/CD Pipelines, Model Validation, MLOps Concepts, Jira, ServiceNow, Agile (Scrum), Docker.

## PROJECTS AND RESEARCH

- 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 leveraging ARIMA models 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*
- Reddit Community Behavioral and Trend Analysis** Sep 2023
- Analyzed political subreddit activity by implementing LSTM and BERT models for sentiment classification and topic modeling.
  - Delivered 87% classification accuracy and deployed a modular ML Ops pipeline for scalable experimentation.
  - Deep Learning, LSTM, BERT, TensorFlow, Scikit-learn, NumPy, ML Ops*

## 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, Smart energy assistant
- Gandhi Institute of Technology and Management - GITAM University** | *Visakhapatnam, India* Jun 2017 - Jun 2021
- Bachelor of Technology in Computer Science and Engineering*
- Relevant Courses:** Database Management, Artificial Intelligence, Machine Learning, Probability & Statistics, Cloud Computing