

Akshitha Kumbam

+1 (650) 405-9026 • New York City, New York

ak11071@nyu.edu • [LinkedIn](#) • [Medium blog - Akshitha Kumbam](#)

EDUCATION

Master of Data Science, New York University, Center for Data Science

Aug 2023 - May 2025

Bachelor of Mechanical Engineering, Minor in Data Science, Birla Institute of Technology and Science (BITS)

Aug 2018 - May 2022

EXPERIENCE

AMAZON WEB SERVICES

Solutions architect intern ML | **AWS**, **Streamlit**, **LLM**, Prompt engineering, Linux, RAG, GenAI

Cupertino, California

May 2024 - Aug 2024

- Built an AI-powered code generation tool leveraging AWS Bedrock and Streamlit, reducing manual effort through automated code validation and optimization using EC2.
- Guaranteed code correctness using iterative feedback loops, chain-thought prompt engineering, and dynamic prompts based on inputs.
- Integrated a Retrieval-Augmented Generation (RAG) system using AWS Kendra to retrieve, rank, and leverage relevant documents for precise and context-specific response generation.

ELTROPY INC

AI Backend Intern | **Transformers**, **FastAPI**, JavaScript, **Docker**, Kubernetes, Encoder-Decoder models

Milpitas, California

Feb 2024 - May 2024

- Developed a deep learning pipeline using BERT and SentenceTransformers to quantify semantic similarity between sentences using embeddings. Fine-tuned SBERT for improved performance, utilizing cosine similarity and error metrics (MSE, Pearson) to validate model accuracy.
- Acquired expertise in transformer models (LLaMA, Mixtral, GPT) and text vectorization/embedding techniques, enabling effective text data representation in ML models.

Data Engineer | Python, SQL, Cloud - **AWS**, **Airflow**, Redshift

Sep 2022 - Jul 2023

- Built pipelines to automate periodic data ingestion to Data lake using cloud technologies (ETL, Athena, Lambda, S3, CloudWatch, Databrew).
- Created monetizable analytics dashboards on AWS Quicksight end-to-end.

BLOGGER - [Medium blog](#), [Substack blog](#)

Jan 2024 - Present

Published articles on AI/ML topics, including transformer architecture, large language models, retrieval-augmented generation, and Guardrails.

Currently doing [AI inshort series](#).

WALMART LABS

Data Science Intern | **Pandas**, Scikit-learn, **Tableau**, NumPy, Matplotlib, Seaborn, Data visualization

Bangalore, India

Jan 2020 - Aug 2020

- Uncovered high-value customer segments by applying K-means clustering, identifying a key group with a higher purchase frequency.
- Visualized market trends and attribute relationships using Tableau, providing actionable insights for business stakeholders.

SKILLS AND COURSES

Technical Machine learning (ML), Deep learning, Natural language processing (NLP), Data structures and algorithms, Computer Science, Computer vision, Artificial Intelligence (AI), Multimodal learning, Object-oriented Programming, Data mining, Operating systems

Languages Python, Java, C++, SQL, C

Tools AWS, Git, Apache Kafka, Hadoop, MapReduce, MySQL, Spark

PROJECTS - [Finetuned gemma](#), [Custom-curated dataset](#)

Finetuning Gemma-2-9B - Unlocking Creative Storytelling | PyTorch, **Unsloth**, **LoRA**, GPU optimization

- Fine-tuned Gemma 2 to build a storytelling model, generating engaging and child-friendly tales from user-provided opening lines.
- Developed a customized dataset(1k) with AI-generated one-liners for short stories, ensuring age-appropriate themes.

Seq2Seq Neural Machine Translation | **PyTorch**, Docker, Hugging Face **Transformers**, RESTful API

- Fine-tuned T5-base model using learning rate scheduling and gradient accumulation to achieve BLEU score: 53.2.
- Designed a data pipeline to preprocess French-English sentence pairs from diverse datasets (TED Talks, WMT14).

PUBLICATIONS AND AWARDS

- LIDS : Lightweight Dynamic Scheduling Technique for 6G-enabled Massive LoRa based IoT Systems. Published by IEEE GLOBECOM 2022 conference, Rio de Janeiro, Brazil - Dec 4, 2022.
- Active member of Association for Computing Machinery-Women (ACM-W) (Link to my page). ACM-W award for attendance at IEEE GLOBECOM 2022 sponsored by Google through ACM-W scholarship program.