

# Saksham Adhikari

<https://www.linkedin.com/in/adhsaksham/> | 737-315-1963 | [pqo14@txstate.edu](mailto:pqo14@txstate.edu) | <https://github.com/Tar-ive>

## Education

### Texas State University (TXST)

Bachelor of Science in Computer Information Systems GPA 4.0/4.0

2023 - Current

San Marcos, TX

## Professional Experience

### Google

August 2025 - Current

#### TPU Cloud Student Researcher

Remote

- Leading an advanced TPU optimization research on Google Cloud TPU v6e architecture with contributions to the open-source vllm project, implementing architecture-adaptive attention backends for large language model inference on next-generation hardware
- Architected multi-agent reinforcement learning system for power grid intelligence using 618-dimensional state space and 145-dimensional action space, implementing hierarchical control with Strategic, Operational, and Safety agents coordinated through 8-head multi-attention mechanisms on TPU v4 clusters

### Center of Analytics and Data Science @ TXST

July 2025 - Current

#### Student Programmer

San Marcos, TX

- Engineered a full-stack [researcher discovery tool](#) using UMAP, semantic clustering, and LLMs to process 2,454 academic papers into a dynamic 2D map for the **ExpandAI grant**.
- Optimized the data payload to 259KB, achieving **sub-500ms load times** to enable seamless discovery of interdisciplinary connections.
- Architected a full-scale **DevOps lifecycle** with a CI/CD pipeline using **GitHub Actions** and **Vercel**, authoring comprehensive unit tests via **TDD** to ensure production stability.
- Integrated **Sentry** for real-time error monitoring and performance analysis, ensuring high availability for all users.

### Texas State University

May 2025 - July 2025

#### Software Development Intern

Austin, TX

- Developed **agentic AI systems for healthcare applications** using GAIA framework, deploying a high-performance inference gateway (llama-nexus) integrated with **Qdrant** vector database for semantic search and **TiDB** for distributed patient data persistence
- Optimized **autonomous agent workflows** in Llama Edge gateway server, implementing and testing tool-use capabilities and function calling across multiple LLM models to enable reliable healthcare decision support systems
- Executed comprehensive testing protocols to evaluate agentic AI system reliability in healthcare contexts, discovering **1 critical SQL injection vulnerability** while validating agent behavior, data flow integrity, and system performance across multiple test scenarios.

### Translational Health Research Center

November 2024 - Current

#### Undergraduate Research Assistant

San Marcos, TX

- Engineered a **HIPAA-compliant Python NLP pipeline** to perform sentiment analysis on 619 patient interactions, securely uncovering insights into mental health for a grant-funded AI research project.
- Reduced data processing time by **40%** by architecting a scalable data pipeline using **Pandas** and **PostgreSQL** to manage 30k+ HIPAA-compliant patient data points.
- Developed and executed the entire analytical pipeline on a cost-optimized **AWS SageMaker** instance, delivering the project **200% under budget** (< \$150 total cost).

## Publications & Research

**Co-Author**, “Exploring rural women’s healthcare access through social vulnerability profiles: A cluster analysis of regional survey data in Texas”

- Developed a novel vulnerability profiling framework using unsupervised machine learning that identified 7 distinct subgroups of rural women; demonstrated that these AI-derived profiles were stronger predictors of healthcare access than traditional demographic factors like race or insurance status.
- Engineered and deployed an end-to-end clustering pipeline in Python (scikit-learn) by systematically evaluating four algorithms (K-means, Hierarchical, GMM, Spectral) to identify the most robust model, achieving exceptional cluster validation scores (Silhouette Score: 0.93, ARI: 1.0).
- Executed advanced statistical validation using a Generalized Linear Model (GLM) in Python to confirm cluster significance, after performing dimensionality reduction by engineering composite binary features to enhance model stability and interpretability.
- Integrated and visualized multidimensional data by developing geospatial maps of social determinants with ArcGIS Pro and generating 3D scatter plots in Python to illustrate cluster separation and demographic distributions.

**Lead-Author**, “QuantaFold: Scaling Protein Language Model Fine-tuning to 5000 Families Through Systematic Optimization”

- Pioneered a systematic HPC optimization [pipeline](#) that transformed training failures (19+ hour projected) into more optimized training runs for protein language models, achieving 78% training time reduction while scaling [ESM-2](#) (Meta AI) fine-tuning to 400,000 sequences across 5,000 protein families using the [pfam](#) dataset released by Google.
- Developed an intelligent data curation strategy using stratified sampling that reduced dataset size by 3.3x (1.34M → 400K sequences) while preserving statistical diversity across all protein families, enabling feasible training without sacrificing biological representation or model generalizability.

- Validated optimization impact through comprehensive W&B monitoring, demonstrating 97.9% accuracy on specialist tasks (1.26 hours) and 60.32% accuracy on generalist classification (4.17 hours), with production deployment achieving 63.914 samples/second inference throughput; research findings selected for poster presentation at SC25 (International Conference for High Performance Computing).

## Extracurricular Experience

---

### ACM AI @ TXST Vice President

December 2024 - Current  
San Marcos, TX

- Lead a 70+ member club in building full-stack applications deployed on **AWS EC2 Linux instances**, including a campus marketplace with a **React frontend, Node.js & PostgreSQL backend**, and **Tailwind CSS styling**
- Led a 4 person team in engineering and deploying a ML professor recommendation bot, deployed **FastAPI** backend for suggestions using **GCP CloudRun** using **LightFM, scikit-learn** and **Nltk** library for the club discord channel, achieving an **87%** positive feedback rate across **484** interactions
- Organized a 50 person [AI ethics debate](#) across faculty, industry professionals and local Texas government employees and received \$1.7k in funding via sponsorships for the event.

### AI4All Ignite

September 2024 - February 2025  
Remote

### Machine Learning Fellow

- Led a 5 person team virtually and trained a SVM model to predict early signs of Alzheimer's using the Darwin dataset over a 6 month AI accelerated fellowship program
- Developed an [app](#) to make Alzheimer's early detection technology accessible to everyone, and presented a poster highlighting 91% prediction accuracy in a Research Symposium.

### BokoHacks 2025 Organizing Committee

January 2025 - May 2025  
San Marcos, TX

### Student Volunteer

- Helped facilitate communication and on ground transport for hackathon judges, and successfully organized a 100+ person hackathon

## Projects

---

### [Grants-MCP](#) (2025)

- Architected production-grade Model Context Protocol (MCP) ecosystem with Python and TypeScript, building comprehensive government grant discovery platform for AI assistants with 3 separate applications (core server, web UI, marketing site)
- Engineered real-time API integration replacing mock data with live Simpler Grants API calls through async-to-sync bridge architecture, enabling AI assistants to query 180k+ government grants with pagination support and formatted responses
- Achieved **389+** downloads on PulseMCP platform and gained visibility across multiple MCP developer forums, establishing the tool as a go-to resource for developers integrating government grant search capabilities into AI workflows with 14 GitHub stars + 3 forks

### [Find&Fund](#) (2025)

- Developed a proposal coach using Python and a finetuned Llama model, resulting in 3 successful in-state grant awards for PhD students in a 14-person pilot group
- Architected and deployed a Flask API with a /analyze endpoint that uses LlamaIndex to parse uploaded grant PDFs, providing targeted funding recommendations by analyzing "Specific Aims" statements.
- Generated tailored proposal feedback by fine-tuning an LLM using Low-Rank Adaptation (LoRA) on 70+ grant proposals, creating lightweight, efficient adapters to guide users on successful proposal structure.

## Awards

---

### • TXST Datathon – 1st Place (2025)

Developed a greedy parking space optimizer and ML algorithm using **Tensorflow** to strategically manage parking spots

### • Novo Hacks – Best Design (2024)

Built a **foodbank app** leveraging AI agents to maximize food displacement strategies for underserved users

### • National Economics Olympiad - Gold Medalist in Business Case Analysis - Kathmandu, Nepal (2022)

– 2022: Presented a business plan to sell Himalayan fresh water

## Scholarships

---

### • Texas State Merit Scholar (2023)

Full-tuition scholarship awarded to 15 students schoolwide for outstanding leadership, academic excellence, and commitment to social impact

### • AKAEF Undergraduate Launch Scholar (2024)

\$5,000 scholarship awarded for technical potential and commitment to creating a more inclusive and equitable tech industry

### • Merry Kone FitzPatrick Endowment Scholarship Recipient (2024)

\$4,000 scholarship awarded for honors excellence and commitment to creating a more inclusive and equitable campus community

### • Montgomery Endowment Web Service Scholarship Recipient (2024)

\$2,000 scholarship for excellence in web service, leadership in housing communities, and passion for change

## Skills

---

Python | JavaScript | JAX | Rust | SQL | React | HTML | CSS | Git | Linux | Terraform | Tensorflow | OpenCV | Swift | Node.js | NLP |