

Akhil Theerthala

Bengaluru | akhiltvsn@gmail.com | 8179528501 | LinkedIn | GitHub | HuggingFace

Summary

Senior Data Scientist with 2 years shaping large-language model products: Curating reasoning datasets, fine-tuning language models, and shipping quantized models that run on a single GPU. Currently interested in building domain-specific reasoning models to deliver in low-latency, cost-efficient environments.

Experience

Research Affiliate, Financial Services Innovation Lab, Georgia Tech August 2025 – Present

- Curating US finance corpora and generating high-quality financial QA pairs with explicit chain-of-thought.
- Developing generalized financial reasoning models and running ablation studies across dataset-difficulty bands and model scales.

Senior Member Data Scientist, Perfios Software Solutions April 2025 – Present

- Fine-tuned vision language models to address table structure recognition in production, achieving an average TEDS score of 0.85 on mixed bank statements.
- Developed a framework for document decipherability that decreases the time required to analyze documents with unreadable text by approximately 20%.
- Developed agent frameworks for underwriting, using Python, PyTorch, and NLP techniques to demonstrate product workflow concepts.

Member Data Scientist, Perfios Software Solutions June 2023 – April 2025

- Implemented and optimized a multimodal document classifier, reducing end-to-end processing time from 8 seconds to 200 ms.
- Engineered the table detection module by evaluating different architectures, achieving a 27.6% increase in detection performance.
- Developed a transformer-based row detection module for table structure recognition, enabling document input in 40 ms.

Projects

Kuvera - A Single-GPU Personal Finance Advisor HuggingFace Dataset

- Trained 8B and 14B-parameter language models on a single GPU for on-device personal finance advice, reaching 3,000+ downloads on Hugging Face in one month..
- The further versions of the model have been published in a private repository, and are planned to be released along with the corresponding paper.

Personal Finance CoTR Dataset HuggingFace Dataset

- Generated 20,000 CoT samples using a cost-effective pipeline and achieved over 1,000 downloads.
- Engineered a multistep generation pipeline, leveraging DeepSeek-V3-0324 outputs to guide responses under budget constraints for data scraping and generation.

Social Media Engagement Drivers for CrowdFunding projects GitHub

- Designed and implemented an NLP pipeline to transform unstructured social media text into predictive features for machine learning models.
- Achieved an 0.82 F1 score with ensemble methods and evaluated 50+ variables, including engagement metrics, topic scores, and text complexity
- Analyzed partial dependence plots to quantify the relationship between social media attributes and crowdfunding success.

Education

Indian Institute of Technology, Kharagpur, B. Tech in Aerospace Engineering Aug 2019 – May 2023

- **Related Coursework:**

Graphical and Generative Modelling for ML | Dependable and Secure AI-ML | DL Foundations and Applications | ML Foundations and Applications | AI Foundations and Applications

Publications

A Data-Centric Framework for Training Behaviour-Aware Personal Finance Language Models

FinNLP workshop, EMNLP 2025

Certifications

Generative AI:

| | |
|------------------------------------------------------------------------------------|------------|
| The Reasoning Course , <i>HuggingFace</i> | April 2025 |
| Generative AI with Large Language Models , <i>DeepLearning.ai, Coursera</i> | Feb 2025 |
| Generative AI Nanodegree , <i>Udacity</i> | Dec 2024 |

Machine Learning:

| | |
|--------------------------------------------------------------------------|----------|
| Machine Learning in Production , <i>DeepLearning.ai, Coursera</i> | Jun 2023 |
| Deep Learning Nanodegree , <i>Udacity</i> | Apr 2022 |
| Deep Learning Specialization , <i>DeepLearning.ai, Coursera</i> | Dec 2021 |

Programming:

| | |
|------------------------------------------------------------------------------------|----------|
| AI Programming With Python Nanodegree , <i>Udacity</i> | Nov 2022 |
| Python Programming Specialization , <i>University of Michigan, Coursera</i> | Jul 2021 |

Skills

Languages: Python, C, SQL

ML Frameworks: PyTorch, TensorFlow, Keras, HuggingFace Transformers, TRL, Scikit-learn, OpenCV, NLTK

Web Scraping & Automation: BeautifulSoup, Selenium, Scrapy

Data Frameworks: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Pillow

Experimentation Tools: Tensorboard, Git, Docker

Achievements

- Achieved 1st Place in the Reasoning Dataset Creation Competition 2025, organized by Bespoke Labs, HuggingFace, and Together.ai.
- Attained finalist position in the Perfios IdeaFest competition for the year 2024-25 with a novel product concept in the field of healthcare.
- Chosen for the Udacity-AWS AI & ML Scholarship, Summer Cohort 2023