Akhil Theerthala

Bengaluru | akhiltvsn@gmail.com | 8179528501 | Linkedin | Portfolio | HuggingFace

Summary

Applied Scientist with 2.5+ years of hands-on experience specializing in Natural Language Processing (NLP) and Vision-Language Models (VLMs) for complex document intelligence and domain-specific reasoning. Proven technical ownership over the full ML lifecycle, translating cutting-edge research into scalable, deployed production systems. Achieved significant, quantifiable impact, including a 97.5% reduction in inference latency for critical document processing and a 27.6% accuracy boost in generalized table detection.

Professional Experience

Senior Member Data Scientist | *Perfios Software Solutions* Apr 2025 – Present

- Pioneered the application of advanced VLMs (PaliGemma2) adapted via LoRA on domain-specific financial data, establishing a TEDS score baseline of 0.85 in prototype environments.
- Designed and validated the architectural feasibility of multi-stage agentic reasoning workflows (Google-ADK) for complex tasks (underwriting, claims, retention risk), to validate the feasibility of these advancements over the existing internal pipelines.
- Developed a Reference-Free algorithm to quantify document legibility, utilizing a regressive Vision Transformer (ViT). Filtered low-fidelity inputs with **92% precision**, preventing downstream hallucination and reducing compute costs.

Member Data Scientist | *Perfios Software Solutions* Jun 2023 – Apr 2025

- Resolved inference latency bottlenecks ($8s \rightarrow 200ms$), enabling real-time document classification, by compressing a large multimodal classifier into a lightweight student architecture via Distillation and quantization techniques, preserving statistical parity in F1-scores.
- Boosted generalized table detection accuracy by **27.6%** for improving the performance of the TSR module, through the careful, semi-synthetic data curation for finetuning model and rigorous evaluation of YOLOv8 models.
- Operationalized the existing vision-only Table Structure Recognition (TSR) systems w by integrating a semantic row-detection module (via fine-tuned text-encoders), without significant impact on the overall TaT (**<40 ms**).

Additional Experience

Research Volunteer | *Financial Services Innovation Lab, GeorgiaTech* Aug 2025 – Present
(External Collaborator)

- **AAAI 2026 Acceptance:** Co-authored FinForge,a semi-synthetic benchmark generation pipeline for financial tasks. (Accepted at AAAI Workshop on Agentic AI).
- **Resaerch Focus:** Contributing to the **FinGT (Financial Generative Transformers)** project, focusing on enhancing financial domain adaptation and logic adherence in LLMs.

Open Source Contributor | *Hugging Science (AI-for-Food-Allergies)* Oct 2025 - Nov 2025

- Curated high-quality public datasets for food allergy detection to support AI safety in health domains.
- Engineered an interactive **Dataset Explorer**, enabling the research community to easily visualize and analyze allergy data distributions.

Project Experience

Reasoning Dataset Creation Challenge Winner - 1st Place (Global)

- Secured **1st place globally** (out of 150+ teams) in the reasoning datasets challenge hosted by Bespoke Labs, HuggingFace and Together.ai, creating a consolidated synthetic reasoning dataset.
- Engineered a synthetic data generation pipeline for personal finance domain using raw user queries from the web.
- Demonstrated that high-quality synthetic data could allow a small Size, 7B model to rival the reasoning capabilities of larger foundational models, e.g., 14B,24B.

Density Vs Diversity - Data Curation Strategy validation for VLMs Blogpost

- Investigated the trade-offs between dense and diverse sampling strategies by curating 15k-sample synthetic datasets for Vision Language Models (VLMs).

- Conducted comparative analysis of 4B and 8B models across in-domain and RealWorldQA benchmarks to assess reasoning capabilities and OOD generalization.

Kuvera: A Data-Centric Personal Finance LLM HuggingFace Dataset

- Curated a high-quality instruction-tuning dataset specifically for personal finance, filtering for Indian context and regulatory accuracy.
- Fine-tuned 8B/14B parameter models to outperform general-purpose baselines on finance-specific reasoning tasks.
- Open-sourced the full stack (Dataset, LoRA adapters, Quantized weights), driving **35000+ downloads** and fostering community-driven deployment.

PaperStack - Interactive Research Paper Reading Assistant HuggingFace Space

- Architected a research dashboard that transforms PDF manuscripts into interactive, modular summaries, enhancing information retrieval speed.
- Implemented a **tiered inference architecture**: enables dynamic switching between a low-latency "Fast Mode" (for rapid screening) and a "Deep Analysis" mode (for comprehensive summarization / context-heavy query resolution), optimizing user cost-performance ratios.

Causal Analysis of Social Media Signals on Crowdfunding Success GitHub

- Designed a multi-stage experimental framework to decouple the impact of social engagement metrics from fundamental content features on funding rates.
- Conducted stepwise ablation studies to quantify the marginal lift of social signals, demonstrating that content features drive success independently of engagement metrics.

Education

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

Relevant Coursework:

- Graphical & Generative Modelling for ML
- Dependable & Secure AI-ML
- Linear Algebra for AI and ML
- Theories of Language Comprehension
- Machine Learning Foundations
- Financial Analytics

Publications

G. Matlin, **A. Theerthala**, et al. (2025). **FinForge: A Semi-Synthetic Benchmark Generation Framework for Finance.** *Agentic AI in Financial Services, AAAI 2026*.

A. Theerthala (2025). **A Data-Centric Framework for Training Behaviour-Aware Personal Finance Language Models.**
FinNLP, EMNLP 2025.

Certifications

- **The Reasoning Course** (*Hugging Face*)
- **Generative AI with LLMs** (*DeepLearning.AI*)
- **Generative AI Nanodegree** (*Udacity*)
- **Machine Learning in Production** (*Coursera*)
- **Deep Learning Specialization** (*DeepLearning.AI*)

Skills

NLP & LLMs: HuggingFace (Transformers, TRL, PEFT/LoRA), LangChain, vLLM, Prompt Engineering
Deep Learning & ML: PyTorch, TensorFlow, Scikit-learn, OpenCV, NLTK, NumPy, Pandas
Developer Tools & MLOps: Docker, Git, Linux
Visualization: Matplotlib, Seaborn, Plotly, Gradio/Streamlit