AASTIK

8076945930 ◇ aastik.solanki99@gmail.com ◇ Portfolio ◇ linkedin.com/aastik ◇ github.com/aastik

EXPERIENCE

Lead Machine Learning Engineer, *Samsung Research India* *March 2025 - Present*

Norm Adaptive Localized (NormAL) LoRA

- Developed an **adaptive-rank variant** of LoRA that learns and optimizes layer-ranks during training based on importance resulting in up to **50 percent less variance** in multiple benchmarks.
- Introduced 2 novel **Knowledge Concentration Regularization** techniques, that concentrate knowledge within a core subset of LoRA parameters, enabling pruning of the rest without performance loss.
- Achieved up to **37% LoRA adapter size reduction** across various NLU and NLG, enabling efficient deployment on edge devices.

Senior Machine Learning Engineer, *Samsung Research India* *August 2023 - March 2025*

Semantic Search using Vision Language Model, in collaboration with University of Cambridge

- Transitioned from tag-based to natural language search using Vision Language Models (VLM) on edge devices, improving search relevance for **10 million+ flagship devices**.
- Experimented with a novel training regime of subset selection of negative pairs via submodular selection in SIGLIP for image-text alignment pretraining for **11 percent faster training**.
- Proposed new metrics (Recall@80p) for Image retrieval task, and optimized VLM pipelines on edge devices, reducing model size by **16 percent and inference time by 25 ms per inference.**

NEXIN:**N**egative **Ex**clusion **In**ference

- Proposed **NEXIN**, **improving result accuracy by over 85% for** ΔU for exclusion queries.
- Authored and submitted the **NEXIN paper to EMNLP 2024**, covering 100+ hours of research and development, playing a pivotal role as the main contributor and **secured a patent**.

TestSuite

- Developed a model evaluation tool(TestSuite) used by 5+ teams, **reducing evaluation time by 50**%.
- Deployed a tool to evaluate 7 VLM pipelines on **60+ datasets, saving 100+ work-hours/month**.

PROJECTS

LLM based Multi-Agent Medical Assistant (arXiv) *November 2024 - December 2024*

- Demonstrated top performance, **winning** 1st **place** at the LLM Agents Hackathon (Berkeley RDI).
- Built multi agent interaction framework for monitoring 4+ vitals, reminders, and intelligent scheduling.

Know your Tweeters (GitHub) *July 2021 -November 2021*

- Implemented Django-based platform with **87 % accuracy** in detecting bots and black market users on Twitter.
- Automated data collection for **1,500+ accounts** and identified **65 attributes** to identify bots via Twitter API
- Published in ComPE 2021, winning **Session's Best Paper Award**.

EDUCATION

MTech in CSE - Indian Institute of Technology (IIT), Bombay *August 2021 - June 2023*

Specialization: *Machine Learning and Artificial Intelligence* - CPI 9.5

Secured **All India Rank 12** in GATE CSE 2021

SKILLS

Languages: Python, C/C++, Java, Rust, Kotlin

Frameworks: TensorFlow, PyTorch, Keras, Pandas, NumPy, Scikit-learn, ONNX, PyTorch Lightning, MLOps

Others: Django, Flask, LLM finetuning, AI agent modeling, DevOps

PUBLICATIONS & PATENTS

- Published paper **Framework for Co-distillation Driven Federated Learning to Address Class Imbalance in Healthcare** in **CODS-COMAD 2024**.
- Published paper **Machine Learning-Based Identification of Collusive Users in Twitter Stream** in **ComPE 2021** and won the **Session's best paper award**.
- Filed a patent for **NEXIN**, introducing novel metrics and benchmarks for exclusion-based image-text retrieval.