VARUN MEDAPPA

CONTACT

✉ varunfer@gmail.com

📍 New York, NY

Skills

- Languages
    - Python
    - Bash
    - Pytorch
    - Tensorflow
    - SQL
- Platforms
    - AWS
    - Databricks
    - GCP

Accolades

- 2021 Best AI project award at a Vista-organized hackathon
- Patent for Multimodal Classification : https://patents.justia.com/patent/12217493

Education

Bachelor of Engineering, Industrial Engineering
R. V. College Of Engineering

Master of Science, Operations Research
Columbia University

PROFESSIONAL SUMMARY

Data Science Leader with 9+ years delivering production ML across multimodal Audio/Video, applied GenAI, misinformation detection, machine translation, and ads attribution. Hands-on leader and founding member of IAS Multimodal ML team whose algorithms power ~$150M in annual revenue (~30% of company) with 34% YoY growth.

EXPERIENCE

Integral Ad Science | New York, NY | 2019 - Present

Staff Data Scientist - Applied Gen AI Team

- Served as the interim manager of the Innovation team, managing 8 FTE across ML research, ML engineering, Dev Ops, Model deployment strategies.
- Led PEFT/LoRA adoption for A/V deepfake detection, slashing compute ~80% and boosting experiment throughput ~3× without degrading performance.
- Pitched, designed, and shipped an internal python package for multimodal video retrieval and labeling pipeline (text/image/video embeddings + vector index, ANN search) that routes only positive matches to VLM/LLM, cutting production labeling cost and turnaround by 99.5% with high-precision outputs.

Senior Data Scientist - Media Classification Team

- Founding member of the Innovation/Media Team at IAS. Developed a custom Brand Safety Multimodal Video Classification product built using Tensorflow and deployed on ECS Fargate, which processes 40 years of video and 3 trillion impressions per year across multiple social media platforms (TikTok, YouTube, Meta, Twitter). This powers our social suite that brings in 150 million dollars of revenue (30% of company revenue) and is growing YoY at 34% (Patent approved : https://patents.justia.com/patent/12217493)
- Directed the end-to-end development and deployment of a Misinformation Detection initiative, utilizing advanced multimodal and distilbert models extended for longer token length classification to identify false themes in 22 Million+ videos per day

Data Scientist - Media Classification Team

- Developed a pseudo-labeling pipeline that reduced Machine Learning Model Development costs by over 90% across 12 Unsafe Brand Safety and 37 Contextual video categories by fine tuning vision language models using SigLIP and Vicuna.
- Led a team to generate synthetic Twitter data with large language models (GPT-Neo), enhancing our Brand Safety classifier's Precision by 49%.
- Developed custom machine translation models using OpenNMT for 42 language pairs, translating 75 trillion characters annually and reducing costs by 67% of our core contextual classification pipeline

Associate Data Scientist - Incrementality/Lift team

- Supported development of an Online Conversion Lift pipeline which ingested 8 billion impressions per day and attributed the cause and effect on conversion rates of advertisement campaigns

Capillary Technologies | Bangalore | 2016-2017

Business Analyst

Mu Sigma | Bangalore | 2015-2016

Data Scientist