

Samuel Davis

Machine Learning Engineer



+91 8652289114



samuel.davis.aiml@gmail.com



github.com/sam99dave

SKILLS

Language - Python, SQL

Deep Learning Frameworks - PyTorch, Tensorflow

Web Framework - Flask, FastAPI

App Framework - Streamlit

Cloud Vendors - AWS, GCP

Core Skills - NLP, SageMaker, Docker, Information Extraction and Retrieval, Layout Analysis, Transformers

Experience on - OpenCV, Topic Modelling, Distributed Training, Boosting, LLMs, Vector Stores, LangChain, PEFT, GANs

ACHIEVEMENTS

Contributed in setting up sports capability DEMO for AWS Re-Invent which attracted many clients.

Developed critical solutions that fast tracked production deal

CERTIFICATIONS

- Nvidia Fundamentals of Accelerated Computing with CUDA
- AWS Certified Cloud Practitioner
- MLExpert Certificate Of Completion
- Deep Learning Specialization
- Machine Learning Course (Stanford)

ABOUT ME

Experienced in designing and building Machine Learning and Deep Learning solutions, with a passion for developing AI solutions that solve real-world problems.

EXPERIENCE

Machine Learning Engineer
QDox Product Development & POCs

Quantiphi | Mumbai
Aug 2022 - Present

Poof of concepts (4)

- Performed transcript analysis for dataset preparation. Guided a team of taggers throughout the data tagging process.
- Trained and evaluated LayoutLM for layout and entity extraction. Deployed inference endpoint using SageMaker.
- Developed complex custom modular scripts for enhancing and aligning layout to retain transcript structure.
- End to end integration using SQS, AWS Lambda, SageMaker Endpoints and required database updates and metadata storing using DynamoDB and RDS.

Product Development

- Performed R&D and integrated AWS Textract Queries as a feature of the product.
- Undertook R&D, trained and evaluated QA models such as DeBERTav3 and Donut. Containerized training & inference script for model onboarding.
- Led an internal POC, displaying SQL query generation and execution capability using Flan-T5, LangChain SQL agents and AWS RDS as a feature addition exploration.
- Developed and deployed APIs for document comparison using visual & textual features using OpenCV, LayoutLM and LLM embeddings. Optimized the codebase using multi-processing.
- Developed and analyzed performance of Topic modelling solutions using LDA (gensim) and BERTopic. Containerized the solution and deployed using AWS Lambda.
- Developed an Active Learning solution, for optimizing and enhancing data labelling process and model performance boosting development progress for POC
- **GenAI** - Contributed in the development and deployment of RAG system using LLMs, LangChain retriever, Agents and vector stores such as FAISS and opensearch.
- Undertook R&D on prompt engineering, exploring few shot and zero shot using LangChain prompt templates and tools.
- Performed research on SFT and RLHF for LLM finetuning using HuggingFace TRL.

EDUCATION

Bachelor of Engineering - Computer Engineer
GPA: 8.97

July 2017 - June 2021