

SHREYAS SOMASHEKAR

+91 8088-442-178 | sheer.somas@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

WORK EXPERIENCE

AI / ML Engineer: [Examroom.ai](#)

Nov 2023 – May 2024

- Engineered a sophisticated multiple-choice question-answer Generative AI application utilizing **LangChain** and **FastAPI**, incorporating CI/CD pipelines to automate testing, while cutting manual question generation time by 40 hours per month.
- Developed **MLOps pipeline** resulting in an increase of 4x in the number of classes detected and 6x improvement in model performance for computer vision models utilizing **ClearML**, **AWS S3**, **lambda** and **AWS Rekognition**.
- Developed a POC to verify audio transcription with on-screen context during exams by concurrently processing audio stream using **OpenAI Whisper** and video stream with **Mediapipe**.

Machine Learning Engineer: [Lincode.ai](#)

Dec 2021 – Nov 2023

- Conducted rigorous testing and analysis of popular open-source object detection models, delivering actionable insights that drove the strategic deployment of the superior model; improved overall system efficiency by reducing training time by 15%.
- Engineered **MLOps pipelines** for training and inference workflows **resulting in a 4x increment** in the number of training and inference instances with real time reporting of metrics through server-sent events **reducing API calls by 90%**.
- Spearheaded a project to fuse Deflectometry principle and AI solutions with collaborative robots and industrial cameras, producing an advanced inspection solution; this innovation **reduced manual labor time by 50% in production lines**.

Data Scientist: Huawei Technologies

Jun 2018 – Jun 2019

- Curated custom datasets from Wider Face, SCUT-BP500, DEX, SSRNet, and FaceNet to develop an advanced facial beauty prediction model, leveraging attributes such as age, gender, and emotion for enhanced predictive accuracy.
- Improved performance of object detection algorithm by 12% for video summarization and image aesthetics model for EMUI v9.1.
- Implemented image and video super resolution algorithms, network pruning algorithms to achieve real time inference, yielding a 3dB improvement.

Computer Vision and Deep learning Engineer: Cloudadic

Feb 2017 – Jan 2018

- Leveraged Google Cloud Vision SDK to streamline text detection and extraction processes, resulting in a 50% decrease in document processing time and a 95% accuracy rate for over 10,000 documents processed monthly.
- Generated 6.2 million images with text to enhance the development of a proprietary OCR pipeline.

PROJECTS

LLM as a judge. [Blog](#) | [GitHub](#)

- Incorporated Llama-3.1 8B model as a judge to evaluate the performance of pretrained Llama-3 1B and deepseek-r1 through custom prompts with scoring rubrics. Implemented a streamlit based data dashboard to visualize the results.

RAG based WhatsApp Chatbot using Langchain, Pinecone, and Twilio. [Blog](#) | [GitHub](#)

- Built a RAG based WhatsApp chatbot to answers queries from Tesla car manuals and deployed the application on AWS ECS.

LangChain application to transcribe, summarize and translate YouTube videos. [Blog](#) | [GitHub](#)

- Developed LangChain powered Generative AI solution with OpenAI's GPT- 4 to transcribe, summarize, and translate YouTube videos from URLs enabling users to access video content in a more accessible and efficient manner.

Hand pose classification with Mediapipe. [Blog](#) | [GitHub](#)

- Created a TensorFlow-based model that achieved real-time hand pose classification with an inference time of 4ms.

TECHNICAL SKILLS

Machine Learning: TensorFlow, PyTorch, MLflow, Sklearn, Mediapipe, HuggingFace, LangChain, OpenAI, Mistral, Llama.

Cloud Technologies: AWS, GCP, Kubernetes, Terraform, Docker, AWS SageMaker, AWS Rekognition, AWS Lambda.

Programming Languages: Python, Java, C/C++

Databases: SQL, PostgreSQL, Redis, MongoDB, Pinecone, ChromaDB.

Backend: FastAPI, Django, Celery.

CERTIFICATIONS

AWS Certified Cloud Solutions Architect - [Certificate](#)

Deep Learning Specialization – Coursera

Natural Language Processing Specialization – Coursera

EDUCATION

University of Central Florida

Masters in Electrical Engineering

Aug 2014 - May 2016