MAHII VARIAR

Mumbai, India | mahiivariar26@gmail.com | linkedin.com/in/mahii-variar | Website

EDUCATION

Terna Engineering College, University of Mumbai — Navi Mumbai, India
Bachelor of Engineering (B.E.), Computer Engineering — *Jan 2021 – May 2024*
Relevant Coursework: Data Structures; Operating Systems; Discrete Structures and Graph Theory; Artificial Intelligence; Database Management Systems

- Executive Secretary, Computer Society of India (CSI Student Chapter) — led technical events for 500+ students.
- Tech Lead, Higher Education Cell — organised alumni meets for first- and second-year students.

TECHNICAL SKILLS

Machine Learning & AI: CNN, SVM, Random Forest, KNN, MediaPipe, RAG, LangChain, FAISS, Prompt Engineering, AWS Bedrock, OpenAI API
Programming & Data: Python (Pandas, scikit-learn, TensorFlow/Keras), SQL, FastAPI, spaCy, Ollama
Data Analytics & Visualization: Power BI, Excel, Tableau
Cloud & Data Platforms: Azure Data Services, Databricks, Snowflake, Microsoft Fabric
Tools & Applications: Streamlit, Flask, Git, Docker, SharePoint, Power Apps, Power Automate

WORK EXPERIENCE

Capgemini — Navi Mumbai, India
Data Analyst — *Sept 2024 – Aug 2026*

- Built and maintained 5 major Power BI dashboards and approximately 12 supporting reports, working across Databricks, Snowflake, and SQL Server to source and prepare underlying data.
- Contributed to a group proof-of-concept for an LLM-powered solution generation system using OpenAI, LangChain, and AWS Bedrock, designing prompt engineering pipelines to improve response accuracy and reduce hallucinations in generated outputs.
- Progressed from supporting the team lead's client presentations to independently presenting dashboard insights directly to client stakeholders.
- Queried Snowflake using advanced SQL to extract and transform datasets into reporting pipelines supporting multiple ongoing dashboards.
- Cleaned and transformed 10,000+ records across 3+ source systems to improve reporting consistency.

ACADEMIC & INDEPENDENT PROJECTS

MedCloak — Local GenAI Healthcare De-identification Platform | GitHub — *Independent Project*
Python, FastAPI, spaCy, Ollama, LangChain, FAISS

- Built MedCloak, a local healthcare de-identification app using FastAPI, spaCy, and Ollama to protect 10 PHI categories across 3 file formats (TXT, DOCX, PDF).
- Achieved 98.4% exact-span micro-F1 with 744 true positives and 0 false positives on 120 synthetic clinical notes containing 768 labelled identifiers.
- Improved patient-name recall from 88.9% to 94.9% (+6.0 percentage points) by adding local spaCy NLP to the rules-based pipeline.
- Implemented a 3-layer review workflow — automated redaction, local GenAI suggestions, and human correction — with reversible edits and privacy-safe audit exports.

Real-Time Marathi Sign Language Recognizer | GitHub — *Final-Year Major Project*
Python, MediaPipe, OpenCV, Keras

- Led creation of a novel 5,300-image dataset (53 classes: Marathi letters, gestures, common sentences) with a 4-member team, to address the lack of a suitable public Marathi Sign Language dataset; extracted hand landmarks with MediaPipe.
- Extended the dataset to increase variation in hand size, orientation, and colour.
- Personally trained and evaluated CNN, SVM, Random Forest, and KNN models for gesture classification, comparing precision and error; the CNN achieved ~98.38% accuracy.
- Built the user interface for real-time recognition with text and text-to-speech output.

ACHIEVEMENTS

- Recipient, LiFT Scholarship (2022) by Linux Foundation — awarded under the Women in Open Source category.
- Ranked 87th in GirlScript Summer of Code (GSSoC) and 56th in Hacktoberfest, both within the Top 100 Open Source contributors, 2023.
- Advanced to Round 2 of 3 in Flipkart GRiD 4.0, Flipkart's national-level engineering challenge, 2022.

CERTIFICATIONS

Microsoft Certified: Azure AI Fundamentals (AI-901); Azure AI Apps and Agents Developer Associate (AI-103); Azure Data Fundamentals (DP-900); Power BI Data Analyst Associate (PL-300).