

# Thrangborwell Sohlang

Shillong, Meghalaya • +91 6009453430 • thrangsohlang@gmail.com  
github.com/Thrangsohlang • linkedin.com/in/thrangsohlang

## Professional Summary

AI/ML Engineer with hands-on experience delivering production-grade LLM and computer-vision solutions. Skilled in GPT-powered interfaces, real-time object-detection pipelines, and end-to-end RAG systems. Strong foundation in fine-tuning, optimisation, and backend integration with FastAPI & Streamlit.

## Technical Skills

- **ML/AI:** PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, YOLO, DeepSORT, VGG
- **LLM Stack:** OpenAI, RAG pipelines, LangChain, AWS Bedrock, AutoGen, CrewAI
- **Model Optimisation:** ONNX, TensorRT, OpenVINO
- **Vector DBs:** Pinecone, ChromaDB, FAISS
- **Data & Infra:** SQL, BigQuery, Scrapy, Selenium, Git, Docker, REST APIs
- **Deployment:** FastAPI, Streamlit, VLLM, Ollama
- **Frontend:** HTML, JavaScript, CSS

## Projects

### GPT-Powered Data Visualisation Tool

Internal Tool

2024

- Built an interface enabling business users to generate SQL queries and visualisations via natural language with GPT-4 + LangChain agents.
- Integrated with a 20+-table custom schema; added LLM-based evaluator validation and robust retry/error recovery.
- Reduced the query-to-report loop by >60%.

### Khasi ASR System (Wav2Vec 2.0)

2024

- Developed an ASR model for Khasi (~1 h annotated audio).
- Achieved 22% word-error rate on held-out samples; demonstrated viability for low-resource languages.
- Deployed a real-time transcription demo in Streamlit.

### NLP for Personalised Cancer-Treatment Insights

2023

- Trained multiclass classifiers + NER to extract gene–cancer links from PubMed (Kaggle) corpus using NLTK + scikit-learn.
- Reached 93% accuracy; enabled exploration of gene–treatment relationships for precision medicine.

## Professional Experience

### AI Engineer – Sustainability Economics

Bangalore

Mar 2025 – Present

- Evaluated Google Document AI, AWS Data Automation & Azure Document Intelligence for structured data extraction from PDFs, spreadsheets.
- Designed & prototyped an end-to-end AI-driven ETL pipeline on AWS.
- Achieved 92% extraction accuracy on a manually annotated sample of 20 documents.
- Orchestrated AI Agents using LangGraph, CrewAI & AutoGen and implemented monitoring dashboards (CloudWatch).

### AI/ML Developer – Sarjen Systems (Client of DataLabs)

Ahmedabad

Aug 2024 – Feb 2025

- Enhanced an internal RAG system for document search across PDFs/flowcharts and deployed company-wide.
- Built custom extractors enabling structured search over complex flow-based documents.
- Engineered a real-time golf-ball tracking pipeline (POC) with YOLO + DeepSORT; ~1 min latency on 30-s 120 FPS footage.
- Conducted R&D and fused cutting-edge deep learning and CV methods from research to enhance golf-ball tracking.

### Data Scientist Associate – DataLabs AI

Hyderabad

Jul 2024 – Feb 2025

- Developed a web-scraping pipeline processing ~20 GB of marketing data with Scrapy + Selenium.
- Implemented data-normalisation & outlier filtering, reducing manual cleaning by 45%.
- Prototyped a GPT assistant for structured sales-data queries.

### Data Science Intern – Response Informatics Ltd.

Hyderabad

Oct 2023 – Jun 2024

- Co-developed an internal NLP dashboard that automated SQL generation with live visualisation in Streamlit.
- Conducted EDA and model experiments to support stakeholder analytics workflows.

## Education

---

### M.Sc. Physics – IIT Hyderabad

Aug 2022

- **Project:** modelled gamma-ray-burst afterglows; AIC/BIC favoured broken power-law.
- **Key courses:** Computational Physics, Data Science for Astronomy.

### B.Sc. Physics – St. Anthony's College, Shillong

Oct 2020

## Awards & Languages

---

- **Scholarship:** Inspire Scholarship for Higher Education (2017–2022)
- **Languages:** Khasi (native) · English (fluent)