Vaibhav Singh

Data Scientist/AI Engineer

+919560720582 | vvishus1717@gmail.com | LinkedIn | GitHub | Website | LeetCode

SUMMARY

Results-driven Data Science Intern with proven expertise in architecting AI-driven chatbots, deploying scalable APIs, and engineering end-to-end machine learning pipelines. **Proficient in Python, PyTorch, FastAPI, and cutting-edge NLP/speech technologies** such as BERT, NVIDIA NeMo, and Pyannote. Adept at real-time application development with a robust foundation in data manipulation, machine learning, and web engineering. **Specialized in integrating LLMs and prompt engineering to drive impactful solutions**. Demonstrated excellence through high-impact internship, with a track record of effectively communicating complex technical solutions.

WORK EXPERIENCE

AI Intern, BASAL AI, Bengaluru, India (Remote) **Jun 2025 — Present**

(Agentic AI, OpenAI, Langchain, LangGraph, Pandas ,etc)

- Spearheaded the development of a Proof-of-Concept (POC) for an autonomous AI Sales Development Representative (SDR), defining the complete **Agentic AI** architecture, workflow, and technology stack to automate sales prospecting.
- Engineered a robust data pipeline by integrating multiple lead generation and enrichment APIs, including Apollo.io, Lusha, and Coresignal, to create a unified and comprehensive lead database.
- Designed the foundational framework for the AI SDR, establishing a scalable solution to enhance lead quality, accelerate the sales cycle, and increase outreach effectiveness.

Data Science Intern, OriServe, Noida, India **Jul 2024 — May 2025**

(Python, PyTorch, FastAPI, LLMs, Prompting, etc)

- Collaborated on AI-driven chatbot development and automation solutions, leveraging models like **BERT**, **Pyannote**, **NVIDIA NeMo**, etc and frameworks like **Pandas**, **Pytorch**, **Plotly**, etc to enhance user engagement and operational efficiency by ~20%.
- Designed and deployed scalable **FastAPI** services to integrate AI and streaming functionalities, ensuring optimized system performance and seamless user experiences
- Designed robust **pipelines** for **TTS**(text-to-speech) and **STT**(speech-to-text) datasets, enabling efficient preprocessing, training, and evaluation with advanced speech techniques.
- Applied **prompt engineering** techniques to enhance response quality across varied use-cases, while managing a **NER-based voice chatbot** tailored for a bar and restaurant setting, gaining hands-on experience in **LLM response optimization**.
- Authored an article on Medium: ***Building High-Fidelity Datasets for Superior Text-to-Speech Fine-Tuning***, offering insights on best practices, challenges, and strategies for ensuring superior audio output quality.
- Engineered and optimized PyTorch scripts to **fine-tune** and train diverse ML/DL and Transformer-based models such as **BERT, XTTSv2 (VQ-VAE, GPT-2), Whisper**, etc.
- **Evaluated** and **benchmarked** multiple open-source ASR(NeMo, pyannote) and TTS(XTTS, Orpheus-TTS) models and platforms(Deepgram, Sarvam.ai, ElevenLabs) for performance, accuracy, and deployment feasibility on .

Summer Intern, Directorate of Research, Innovation and Development (DRID), JIIT, Noida **May 2024 — Jul 2024**

(OpenCV, cvzone, PyTorch Lightning, Pygame, NumPy, scikit-learn)

- Engineered an accessible Connect Four game controlled exclusively through American Sign Language (ASL) hand gestures, enabling players with speech or motor impairments to enjoy gaming without traditional input devices.
- Implemented real-time **computer vision** tracking utilizing **OpenCV** and cvzone's HandTrackingModule to detect, isolate, and process hand positions from webcam input with optimized performance on consumer hardware.
- Designed and trained a custom CNN model achieving **99%** accuracy in **PyTorch Lightning** to classify ASL hand gestures, orchestrating multiple convolutional layers to enable precise recognition of "A", "B", and "C" signs as game controls.

EDUCATION

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, Noida, India - *B.Tech, Computer Science & Engineering* **Sep 2021 — 2025**

PROJECTS

HiringAgent | AI-Powered Technical Interview System, Link

- **Implemented dynamic, real-time question generation** based on job profiles and a comprehensive evaluation engine to deliver personalized, data-driven candidate assessments.
- **Architected a resilient, dual-LLM automated interview system (Gemini primary)** to conduct adaptive technical screenings, ensuring high availability and improving hiring consistency.
- **Engineered a multi-phase interview flow with robust integrity and session controls** (e.g., anti-cheating, time/interaction limits) to ensure fair, secure, and standardized evaluations.

Hotel Booking RAG-Based Query System

- Built a RAG system using **SQLite**, **ChromaDB**, and **LLMs (LLaMA2/Gemini)** for natural language queries over hotel booking data.
- Implemented **semantic search** on custom analytics using vector embeddings and integrated LLM-powered query responses.
- Exposed functionality via **FastAPI endpoints** (`/ask`, `/analytics`) for real-time insights and dashboard integration.

PYTONE – REAL-TIME SPEECH EMOTION RECOGNITION

- Developed emotion recognition system using real-time speech and text analysis via ML models to enhance interaction quality.

BCI Games–Brain − ControlledInterfaceGame

- Designed and built a brain-computer interface using EEG signals, enabling real-time in-game actions through smile detection
- Achieved 95% command execution accuracy, reducing cognitive load and enhancing user experience.

AID FOR SPEECH AND HEARING IMPAIRED

- Developed real-time ISL translator using speech-to-text and SIGML-based animations for inclusive communication support.

SKILLS

Programming Languages	:	Python, C++, SQL, C, Html, CSS, Javascript
Data Manipulation	:	Pandas, NumPy, Matplotlib, SQLite
Web Development	:	FastAPI, Flask, Django, Websockets
Machine Learning	:	PyTorch, PyTorch Lightning, tensorflow, scikit-learn, OpenCV
Other Tools	:	Shell, PowerBI