DASHPREET SINGH

Python, AI, ML, GEN AI & LLM Engineer

Linkedin Github dashpreetsinghofficial@gmail.com +91 9560762853

Delhi, India

Professional Summary

Machine Learning Engineer with over 2 years of experience in designing and developing scalable applications using **ML, AI,** and **Generative AI** tools. Proficient in building machine learning models, implementing solutions with Generative AI and Large Language Models (**LLMs**), and creating efficient backend services. Skilled in optimizing database performance, deploying robust CI/CD pipelines, and managing infrastructure using Docker and **AWS**. Known for strong problem-solving abilities and a proven track record of delivering impactful solutions across e-commerce and real-time systems. Driven by a deep interest in **R&D**, staying at the forefront of AI/ML advancements, and consistently delivering high-quality, value-adding outcomes.

TECHNICAL SKILLS

Languages: Python, DSA, C, C++.
Backend: Python, JWT, OAuth 2.0, Flask.
Python Packages and Frameworks: TensorFlow, Keras, Scikit-Learn, GPT, LLM, Natural Language Processing (NLP), Open AI, LLM, Langchain, NumPy, Pandas, Flask, SciPy, Matplotlib, Scikit- Learning, Yolo, Open CV, Torch Vision, Pytorch, Tensorflow Zoo, OCR, Deep Face, Keras, DLIB, Jupyter Notebook, Google Collab, GRPC, Regex, Web Scraping, Selenium, Unit Testing, Supervised Learning, Unsupervised Learning.
Cloud and Deployment: AWS, CI/CD, Docker, Git, GitHub, Pipelines, Jira, Microservices.
ML Architectures: Skilled in designing architectures and implementing supervised and unsupervised ML algorithms, NLP techniques, LLMs and GenAI.
Tools: Visual Studio, Pycharm, Excel, Postman, Tableau, PowerBI, BitBucket.
Databases: SQL, VectorDB.

WORK EXPERIENCE

Machine Learning Engineer — **Contract**
Emcode Technology LLC, Dubai Remote — January 2024 - September 2024

- **LLM Voice Assistant:** using speech recognition, Pyttsx, OpenAI LLM, gTTS and **OpenAI** Whisper, a script for an LLM voice assistant is prepared to provide answers and information in English and Arabic. Additionally, a training script for **fine-tuning** OpenAI LLM generated responses based on custom data that is prepared.

- **Car Destination Prediction:** in this using **data cleaning, feature preparation**, **training KNN** and **Random Forest models** and creating a prediction script to determine the last location of the car based on inputs of time, latitude, longitude and model.

- **Hidden Markov Model for Prediction of Future Location:** In this, the objective is to predict the future location using past data. Initially, a **KNN model** and prediction script are prepared for plotting the vehicle's location on a map. Subsequently, a **hidden markov model** (HMM) is developed to forecast the future location from the current location along with the associated probability

- **Multi-Camera Multi-Vehicle Tracking with AI in a Virtual City:** in this 'Carla' is utilized to generate synthetic data of a virtual city. Following the "Multi-camera vehicle tracking system for AI city article, the electricity (efficient multi-camera vehicle tracking system for intelligent city) method is employed to implement multi-camera vehicle tracking.

Software Developer **Full Time**

Intellypod, USA Remote September 2023 - February 2024

- **LLM CHATBOT:** I designed the **VAG** architecture and I conducted research and development (**R&D**) proof of concept (POC) on OpenAI **LLM**, **google vertex LLM** and **microsoft copilot** to determine the best results among them. Moreover, I created an LLM chatbot capable of reading **PDFs, excel files, text files, web URLs** and storing the data into a knowledge base (**Vector Storage**). LLM then provided answers related to the data. Additionally, I had managed the AI team in terms of support and task assignment. Github

- **HiFi GANs For Speech Synthesis:** I followed an article on **GANs** Speech Synthesis to create synthetic data of LJSpeech. I trained our own model and script to generate synthesized speech.

- **Leads Generator:** I prepared a logic for using **Regex** to find the pattern of **names, address, emails, countries, resorts and vendors** from the docx data. The data then converted into an Excel format and an .exe file was prepared for the client. Github

- **Web Scraping :** I utilized the **Selenium** library and prepared a script for web scraping for the Wyndham Resort website. The scraped data was stored in an SQL database. Additionally, a scheduled script was prepared automatically running the web scraping script on time. Also worked on developing APIs using Flask. Github

Jr. Python Developer **Full Time**

Yoma Technologies, Gurgaon, India September 2022 - September 2023

- **OCR ID Cards Details Extraction:** I trained an **EfficientDet D0 model** (512x512) from **TensorFlow Zoo** on over 3,000 labeled images to detect ID cards and documents. Using **PaddleOCR**, I extracted details from detected documents, including Aadhar, **PAN, driving licenses, GST documents, cheque books, and Voter IDs**.Regex patterns were applied to accurately retrieve specific details, achieving over **90%+ accuracy**. Finally, a **Flask API** was created to output the results in JSON format for easy integration. Github

- **Shelf Monitoring:** I was involved in a shelf monitoring initiative focused on Identifying products such as dabur, zydus, colgate, etc. as well as detecting promotional tags like POSM. My responsibilities included working on the Code for POSM detection, object detection and training the **YOLOv5** model. Moreover, I undertook image annotation for the same dataset. For testing Purposes, I'd also developed a **Flask API**. Additionally, I played a role in managing the labeling team, overseeing data collection, task assignment and coordinated daily activities with the team.

- **Speaker Diarization:** I separated two distinct human voices. I utilized **speech recognition** to detect the language. Then using **OpenAI whisper** to embed and transcribe all the text. Subsequently, We utilized **NVIDIA NEMO** which included **VAD** (voice activity detection Method), **speech detection, speech segmentation, embedding, clustering and speaker labels** to identify the speaker and after separation, Used **NLTK emotion detection** on the caller's text to calculate the review and performance of the caller employee.

Data Science Intern **Full Time**

Technocolab Systems, Indore, India March 2022 - May 2022

- **Voice Assistant :** This voice assistant worked on speech recognition and completed tasks automatically by using libraries **speech recognition, pyttsx3, python and NLP** (Natural Language Processing) Github

- **Image Classification:** This involved predicting the health of crops and plants for farmers using **TensorFlow**, a **CNN** model and Python. Also, created the code from scratch and conducted testing on a **Flask** API. Github

- **Exploratory Data Analysis:** This utilized ML models including the **Smotenn** algorithm to predict the credit **risk level** as either low or high for loans with varying degrees of risk. Github

EDUCATION

Amity University (Online) **Noida, UP, INDIA**

MCA in AI and ML January 2023- January 2025

Maharshi Dayanand University **Jhajjar, Haryana, INDIA**

BCA(Bachelor in Computer Application) August 2019- August 2022

AWARDS AND CERTIFICATIONS

• First Prize in Python Competition - CQS

• Python and Data Science Certification - CQS

• DSA (DATA STRUCTURE AND ALGORITHM) in Python - DUCAT