Parth Sharma

9971361784 | parthshr370@gmail.com | LinkedIn | Twitter | Github

Education

Manipal University Jaipur — Jaipur
Bachelor of Technology in Data Science — *Sep 2022 – Present*

Blue Bells Public School — Gurugram
CBSE — *April. 2007 – May 2022*

Experience

Research Intern — May 2024 – July 2024
Indian Institue of Information Technology Allahabad — *Allahabad, India*

- Developed a fusion model to improve Process Prediction using LSTM and CNN
- The fusion model significantly increased the accuracy along with less training time
- Worked on Fine tunining LLMs for process modelling using POWL amd PM4PY
- Improved the PromoAI Framework of process modelling for large scale databases and detect anamolies in them
- Tech Stack - *Pytorch, Langchain, PM4PY, Pandas*

Projects

HR Portal Agent | *Capstone Project*

- Developed a distributed multi-agent HR system where each agent—built as a self-contained LangChain chain—handles specialized tasks such as OCR-based resume extraction, semantic candidate matching, and dynamic interview question generation.
- Created a multi-agent workflow that outputted custom interview questions generated directly from the candidate's resume, enhancing the personalization and relevance of interview sessions.
- Employed LangGraph to orchestrate agent workflows, enabling real-time visualization, dynamic routing, and on-the-fly adjustments between independent modules.
- Designed the candidate matching agent to combine semantic embeddings, rule-based keyword extraction, and configurable weighted scoring, ensuring high-accuracy candidate ranking.
- Integrated asynchronous communication via RESTful APIs (using Flask/FastAPI) to decouple agents, improve fault tolerance, and optimize processing efficiency.
- Established feedback loops within the interview generation agent to refine questions based on historical candidate data and performance insights.
- Architected the solution for scalability and security by containerizing microservices with Docker, implementing secure API endpoints, encryption, and rate limiting.
- Tools: *LangChain, LangGraph, MongoDB, Flask/FastAPI, Docker.*

Email AI Agent | *Capstone Project*

- Developed a multi-agent orchestration system with LangChain, LangGraph, and LangSmith for managing email workflows, capable of automating the classification and response processes for incoming emails.
- Integrated SMTP server to facilitate real-time email sending, and employed advanced technologies for robust email processing and automatic response drafting based on content classification.
- Designed a comprehensive system architecture that includes specialized agent modules for email filtering, summarization, and response generation, enhancing the efficiency of email management.
- Implemented a dynamic supervisor in LangGraph to coordinate and optimize the workflow between different agents, ensuring effective classification and drafting of replies.
- Enabled the system to automatically draft and send replies to emails based on their classification as spam, urgent, or informational, significantly reducing manual intervention and improving response times.
- Tools: *LangChain, LangGraph, LangSmith, DeepSeek API.*

In-Run Data Shapley | *Research Project*

- Developed a method to compute a Shapley-like contribution score for each training sample with respect to a single validation sample during training on the MNIST dataset.
- Implemented combined training and validation gradient computation within a single forward pass to assess the influence of each training sample on validation loss.

- Utilized first-order approximations through per-sample gradient dot-products to measure the impact of training samples on validation example's loss.
- Conducted multiple backward passes to demonstrate a naive implementation and discussed potential optimizations using forward/backward hooks for efficiency.
- Analyzed and visualized training data influence, facilitating data cleaning by identifying potentially mislabeled or harmful samples.
- Provided extensive documentation and a tutorial blog on Medium, illustrating the practical application and insights from the Shapley values computed during model training.
- Tools: *PyTorch, Librosa, NumPy, Matplotlib.*
- Project code and detailed explanation available on GitHub: Data-Shapley-in-One-Training-Run-Code and Medium: Blog Post

EEG-VAE | *Research Project*

- Developed a custom Variational Autoencoder (VAE) from scratch using PyTorch Lightning for EEG signal analysis
- Implemented comprehensive EEG data preprocessing pipeline, including noise removal, artifact rejection, and feature normalization
- Extracted and analyzed latent representations of EEG signals using the custom VAE architecture
- Leveraged advanced signal processing techniques to transform raw EEG data into meaningful latent space representations
- Utilized PyTorch Lightning for efficient model training and streamlined deep learning workflow
- Tools: *PyTorch, PyTorch Lightning, NumPy, Pandas, SciPy*

Events as Pixels | *Supervisor: O.P. Vyas, IIITA*

- Developed a deep learning model (Fusion Model Architecture - FMA) for predictive process monitoring
- FMA combines LSTM and CNN to capture both temporal and spatial patterns in an event log
- Visualized traces of event logs to a 2D matrix to find unique spatial patterns missed by the temporal methods
- Achieved superior accuracy in predicting the next event and minimal error in timestamp prediction
- Tools: *PM4PY, Keras, TensorFlow, Pandas*

Car Crash Detection using YOLO-V8 | *Minor Project*

- Utilized YOLOv8 for real-time object detection in traffic signal video frames
- Developed a system to plot detected vehicles as data points on a graph for trajectory analysis
- Applied transfer learning techniques to enhance the accuracy of crash detection models
- Employed Pandas and Matplotlib for data manipulation and visualization of detection results
- Tools: *YOLOv8, PyTorch, TensorFlow, Pandas, Matplotlib*

Technical Skills

Languages: Python, Java, SQL, HTML/CSS, C
Frameworks/Technologies: PyTorch, TensorFlow, PM4PY, Librosa, Pandas, NumPy, Matplotlib, PyTorch Lightning, Keras, Flask, Streamlit, EEGLAB
Developer Tools: Git, Docker, Linux, LaTeX, Matlab, EEGLAB, PRoM (Process Mining), GitHub
Data Analysis: Signal Processing, Spectral Analysis, EEG Data Analysis, Statistical Analysis, Machine Learning, Data Visualization
Mathematics: Stochastic Modelling, Linear Algebra, Calculus, Advanced Statistics, Optimization
Research and Documentation: Research Methodology, Technical Writing, Project Documentation, Blogging