APOORV DIXIT

Los Angeles, CA 90007 | 213-574-5136 | apoorvdi@usc.edu | linkedin.com/in/apoorvdixit9

EDUCATION

University of Southern California August 2023-May 2025
Master of Science in Computer Science GPA - 3.73

- Relevant Courses – Analysis of Algorithms, Database Systems, Machine Learning, Natural Language Processing, Operating Systems, Foundations of Data Management, Information Retrieval and Web Search Engines

Pune Institute of Computer Technology August 2017-April 2021
Bachelor of Engineering in Computer Engineering GPA - 9.14

- Relevant Courses – Machine Learning, Artificial Intelligence and Robotics, Data Analytics, Data Mining and Warehousing, Soft Computing & Optimization Algorithms, Web Technology, Software Testing and QA, High Performance Computing, System Programming & Operating Systems, Computer Networks, Computer Organization & Architecture, Discrete Mathematics

EXPERIENCE

JPMorgan Chase & Co. New Jersey, NJ
Quantitative Analyst Intern June 2024-August 2024

- Conducted a Model Review of a fraud detection system in the Corporate and Investment Banking (CIB) Data Science segment to recognize and mitigate potential risks associated with the model, and communicated the identified issues to key stakeholders. Technology Stack: CatBoost
- Developed a benchmarking tool for data extraction models using a two-stage pipeline - the first stage involved text extraction via OCR or PDF parsers, and the second stage utilized LLM for named entity extraction from text. Technology Stack: EasyOCR, PyMuPDF, GPT-4, Llama3.1
- Created a test case generator for stress testing data extraction models by using computer vision and PDF parser libraries to generate noisy versions of input documents, effectively evaluating the robustness of data extraction models. Technology Stack: OpenCV, PyMuPDF
- Formulated a comprehensive study categorizing model-related parameters and concepts into distinct hierarchical level to enhance risk management and prioritization strategies. Prepared a comparison report on Vendor Models in Data Extraction Space
- Worked on the MRGR Model Review Assistant Tool powered by a RAG pipeline with Parent Document Retrievers.

Alzheimer's Therapeutic Research Institute – Keck School of Medicine of USC San Diego, CA (Remote)
Student Intern March 2024-May 2024

- Developed a report dashboard to identify inconsistent and erroneous clinical image data within the electronic data capture (EDC) system. Technology Stack: Django, ReactJS, PostgreSQL, Cypress

Siemens Digital Industries Software Pune, India
Graduate Trainee Engineer → Associate Software Engineer July 2021-July 2023

- Developed and contributed to Zel-X and Xcelerator Share, two cloud-based SaaS platforms facilitating seamless creation, storage, editing, and collaborative sharing of CAD files and projects. Technology Stack: VueJS, ExpressJS, Siemens Web Framework, Jest, BrowserStack
- Collaborated with a Machine Learning Team of 10 members to explore and showcase diverse Artificial Intelligence applications within Computer-Aided Design domain. Filed two **patents** for AI projects (undisclosed)
- Devised an AI pipeline for the conversion of scanned images of CAD drawings into 2D drawing files using Optical Character Recognition, Computer Vision, and Named Entity Recognition. Achieved 95% accuracy in retrieving data from files
- Executed a proof-of-concept for virtual assistant and chatbot functionality within Zel-X using the OpenAI API

BizAmica Software Pvt. Ltd. Pune, India
Machine Learning Intern August 2020-March 2021

- Engineered an AI Pipeline for Bajaj Allianz Life Insurance to identify named entities within scanned legal documents and delineate bounding boxes around extracted data, achieving a 75% accuracy level

Bynry Inc. Pune, India
Web Development Intern May 2020-July 2020

- Assisted in the development of a specialized B2B web application platform catering to 8 clients within the energy and utility sector
- Crafted 20 webpages in Angular crucial for product demonstrations to prospective clients, showcasing the product's capabilities

Center for Development of Advanced Computing Pune, India
Research Intern February 2020-May 2020

- Analyzed the performance and capabilities of Face Embedding Deep Learning Models, such as FaceNet and DR-GAN, using the t-SNE technique
- Benchmarked the time and efficiency of 20 Similarity Search Algorithm Libraries in Python

SKILLS

- **Programming Languages -** JavaScript, Python, Java, C++, R, Solidity, C
- **Web Development -** ReactJs, Django, VueJS, ExpressJS, BrowserStack, Jest, ThreeJS, Angular, Cypress
- **Data Management & Cloud -** PostgreSQL, MongoDB, MySQL, Firebase, Amazon EC2, Hadoop HDFS, DynamoDB, IBM Watson
- **Machine Learning -** Pytorch, Tensorflow, SpaCy, OpenAI, Transformers, Google Vision, Tesseract, OpenCV, Ollama, EasyOCR, LangChain

ACADEMIC PROJECTS

- **Quantized Lag Llama** - Modified the GluonTS Lag Llama model to run on CPU instead of GPU by reducing model size through converting float precision values to integers. Trained and compared ARIMA, SARIMAX, EMA, Meta Prophet, Amazon DeepAR, ARCH, and GARCH models for weather forecasting in Los Angeles from Jan 2014 to Jan 2024
- **Finetuning LLMs using Federated Learning -** Implemented federated machine learning to fine-tune RoBERTa models across all ten datasets of the GLUE benchmark, analyzing performance against standard model variants for comprehensive evaluation
- **Analyzing OSS Ecosystems** – Collaborated with Dr. Alexey Tregubov at Information Sciences Institute to analyze code commits of Open-Source Software using LLMs for malware analysis, topic modeling, summarization, and extraction of code information and authorship styles.
- **Multi-threaded Token Bucket Emulation** - Developed a multi-threaded program to simulate a token bucket traffic shaper for packet transmission, using pthreads for packet arrival, token generation, and server processing. Implemented thread synchronization with mutexes and condition variables, handling both deterministic and trace-driven modes with detailed event logging and statistics. Technology Stack: C, Pthreads, Unix Signals, Mutexes
- **Digi-SOLE Sneaker Store -** Implemented a full-stack e-commerce website for sneakers, at https://digisole-sneakers.onrender.com. Technology Stack: VueJs, ExpressJs, MongoDB, Firebase, Render Hosting Platform
- **Decentralized Vaccine Management System -** Built a supply chain management system for vaccines utilizing Ethereum Blockchain, accommodating four categories of stakeholders. Technology Stack: Solidity, JavaScript, ReactJS, NextJS, Mocha, Ganache
- **Classifier System for Detection of Diabetic Retinopathy -** Created deep learning models to detect the presence and stage of Diabetic Retinopathy in retinal images, achieving 85% precision on test data. Built a web interface using Django
- **Tsunami Prediction** - Developed a machine learning model using Python and IBM Watson to predict the magnitude of probable tsunamis on the Soloviev and Go scale based on earthquake details, utilizing data extracted from the NGDC/WDS Global Historical Tsunami Database
- **PICT Leave Management System** - Worked on a WebApp to facilitate the leave application process at my college. Technology Stack: Linux, Apache, MySQL, PHP
- **MCQ Platform -** Designed the Backend of a WebApp using Django Framework for MCQ-based coding competitions. The project was deployed for Pulzion' 19 - The annual Technical Fest of PICT, serving over 2000 participants

ACHIEVEMENTS AND RECOGNITIONS

- Achieved top three ranks **five** times at Siemens Mainstream Engineering Department Innovation Days
- Commended by Siemens for excellent work in redesigning "Delete" command of Siemens Zel-X
- Honored by Siemens for significant contributions to a new functionality shortly after completing training
- Recognized by Siemens for generating awareness about ChatGPT and Large Language Models among employees
- Secured third prize and the Ethereum track prize at Devcon'21 Hack-Overflow by IEEE NCU Student Branch for prototyping a decentralized competition management system using ReactJS and Solidity
- Awarded High Commendation (Third Place) at the Indian Diplomatic Assembly Model United Nations for representing Nigeria in the Economic and Social Council (ECOSOC) Committee on "African Nations Emerging Out of Conflict"

VOLUNTEERING AND OTHER INITIATIVES

- Volunteered on the assessment panel for the Siemens Scholarship Program and the organizing team for Siemens Family Day
- Invited by USC GRIDS organization to review resumes, helping students enhance their profiles and career prospects
- Conducted multiple training sessions on AI techniques, including OCR, NLP, Computer Vision, and LLM for Siemens employees
- Served as the Head of Marketing and Logistics at PICT Model United Nations Club for three consecutive annual conferences