

Harshad Varma Sagi

+1(704)241-0153 hsagi@ucsd.edu [linkedin](#) [website](#)

EDUCATION

University of California San Diego

Master of Science in Computer Science and Engineering (**GPA: 3.86/4**)

San Diego, CA

Expected: June 2024

Graduate Student Researcher (Data Science), Teaching Assistant (Cognitive Perspectives, Neurobiology)

EXPERIENCE

Graduate Student Researcher

San Diego, CA

Halicioglu Data Science Institute, UC San Diego

Feb 2023 - Present

- Built TutorGPT, an AI-powered tutor designed to address the Bloom 2σ Problem. Leveraged the GPT-4 API, incorporated a student model, and knowledge graphs to provide personalized recommendations, aiming to improve student outcomes.
- Engineered a specialized Piazza bot fine-tuned to answer course-specific questions using retrieval augmented generation. Streamlined student support and reduced instructor workload. (**Tech Stack: Unstructured.io, Cohere, Pinecone, LangChain**)

Graduate Researcher (Software Developer)

San Diego, CA

Cardiac Mechanics Research Group, UC San Diego

September 2023 - January 2024

- Built a data pipeline for biomechanical analysis, including custom data processing and modeling techniques. Developed and deployed a scalable website on AWS for data visualization and analysis (**Tech Stack: ReactJS, NodeJS, S3, MongoDB**).

Machine Learning Intern

Santa Clara, CA

Emagia

June 2023 - September 2023

- Refined the information extraction process for a document processing platform. Achieved a 13% improvement in accuracy in table segmentation across multiple formats using OpenCV for image enhancement and segmentation.
- Deployed a text analysis pipeline using entity recognition capabilities and PaddleOCR for text extraction, resulting in a 25% increase in data interpretation accuracy and error reduction. (**Tech Stack: OpenCV**)

Research Intern

San Diego, CA

UC San Diego Jacobs School of Engineering

June 2021 - Dec 2021

- Developed a self-evaluative ML model for breast histopathology classification achieving 94.4% accuracy. Implemented a novel four-stage solution (Educational Feedback Loop, Adversarial Exploration, Enhanced Learning, Architectural Refinement) and optimized performance using CUDA programming for GPU parallelism. (**Tech Stack: Python, Pytorch, CUDA**)

PROJECTS

ScrapeBot |LangChain, Streamlit, Chroma

- Developed a LangChain based chatbot with a Streamlit GUI, utilizing Retrieval-Augmented Generation (RAG) to combat hallucinations and ensure factual accuracy. Employed OpenAI embeddings for semantic search, Chroma as a vector store, and integrated GPT-4 for advanced language capabilities.

CodeLlama |Llama-2, QLoRa

- Increased Python code generation accuracy by 15% (evaluated using GPT-4) by applying QLoRa fine-tuning techniques to a Llama 7B language model.

LegalLex NER - Natural Language Processing |Python, NLP, Machine Learning

- Developed NER system for legal text entities. Utilized Few-shot learning with prompt engineering on LLMs: RoBERTa, ELECTRA, DeBERTa, GPT-3.5. Achieved an F1 score of 89.80 for GPT-3.5 one-shot learning.

Fault-Tolerant Distributed File Storage System |Go, gRPC

- Implemented a distributed file storage system similar to Dropbox, with support for concurrent read/write accesses from multiple clients, and CRUD/Sync operations on the files. Integrated the RAFT consensus protocol to make the system resilient to the failure of 50% of the servers.

TECHNICAL SKILLS AND COURSES

Languages, Libraries and Tools: Python, Java, C++, Pytorch, GoLang, SQL, OpenCV, Tensorflow.

Coursework: Algorithms, Recommender Systems and Web mining, Probabilistic Reasoning and Learning, Convex Optimization, Distributed Systems, Search and Optimization, Advanced Computer Vision, NLP, Robotics, Computer Architecture.