SHREAYAN CHAUDHARY | shreayan98c@gmail.com | +1 (551) 339-7724 | GitHub | LinkedIn | Blog

EDUCATION

Johns Hopkins University - Master of Science in Computer Science **Expected Spring 2024 (USA)**
CGPA: 3.97/4 | RA: ARCADE Lab, CLSP Lab | TA: Database, Software Engg, Information Retrieval
SRM University - Bachelor of Technology in Software Engineering **Graduated 2020 (India)**

SKILLS (PROFESSIONAL PROFICIENCY)

Languages: Python, SQL, R, Java, JavaScript, HTML+CSS
Frameworks: PyTorch, Tensorflow, JAX, Numpy, sklearn, LangChain, Streamlit, Django, Flask, SpaCy, Huggingface, vLLM
Software Tools: Dask, Spark, MapReduce, Airflow, Ray; MongoDB, PostgreSQL; AWS, GCP, Azure; Git; Docker

PROFESSIONAL EXPERIENCE

Seagate - *Machine Learning Intern* **Jun 2023 - Present (USA)**

- Leveraged GPT-4, Falcon, and Llama2 LLMs for text2SQL for database querying using RAG + chain of thoughts
- Reduced cost by ~67% by finetuning LLMs using PEFT and LORA to optimize the generated SQL, and vector caching
- Extended support to unstructured data (PDF, PPT, DOC, MP4, TXT files) and enabled complex metric computations
- Diminished need for Tableau dashboards by data analyst team savings labor hours and licensing fees by ~85%

Razorthink - *Machine Learning Engineer* **Jan 2021 - Jul 2022 (India)**

- Automated scanned invoices' OCR and Information Extraction by building a rule-based AI with ~93% F1 score and reduced processing time by ~85%
- Developed PyTorch and Tensorflow APIs for no code ML platform by procedurally generating code, used by 40+ firms
- Designed and implemented a scalable data pipeline for Data Ingestion, Wrangling, ETL, and Feature Engineering, handling 1.5+ GB/day

Indian Institute of Technology Bombay - *Machine Learning Research Intern* **Dec 2019 - Jan 2021 (India)**

- Transformed Indian Navy's Russian-English translation process, automating domain-specific docs' translation building a Bi-LSTM Transformer achieving 0.29 BLEU score, reducing human involvement by ~80%
- Pro bono engineered a feedback summarizer using BERT for two NGOs, reducing processing time for 1.3M+ users, increasing resolution rate for 3.7M+ complaints by ~94%

Spocto LLP - *Intern Data Scientist* **Jun 2019 - Jul 2019 (India)**

- Conducted Statistical Analysis & EDA on banks' loan data to visualize customers' propensity and segment defaulters
- Devised and automated an SVM model, reducing staff hours in identifying potential loan defaulters by 70%

Vakrangee Software - *Intern Software Engineer* **Jun 2018 - Jul 2018 (India)**

- Programmed a JSP-based scalable web app for A/B testing with emails, SMS, and push notifications to 1.2M+ users

WisOpt - *Intern Web Developer* **Jan 2018 - Jun 2018 (India)**

- Built analytics dashboard for SRM University serving 600+ profs and 15k+ students for all official communications

PUBLICATIONS (RESEARCH INTERESTS: NLP AND RECOMMENDER SYSTEMS)

- Take a Shot! Natural Language Control of Robotic X-ray Systems for Image-guided Surgery (2024), IPCAI'24 | PubMed
- Recommendation System for Establishing New Businesses using Geospatial Clustering for Multiple Reference Points (2020), NCAIIIP '20; Patented under SRM University | Github | Demo video
- Ensemble Recommendation System using a hybrid decision level fusion of Popularity Model and Collaborative Filtering (2020), ICAIECES'20 | ResearchGate

PERSONAL PROJECTS

PROJECTS LIST | KAGGLE | GITHUB

LookoutX - Smart assistant for the visually impaired using multimodal (speech, text, image) AI | Github | Demo video

- Engineered prototype glasses with camera and mic, and finetuned OpenFlamingo on real-time video to answer users' audio-based questions (Whisper for ASR) conditioned on the image (CLIP ViT for images and LLaMA for text)

OptiRAG - Enhancing Question Answering task with RAG LLMs inference optimization | Github

- Distributedly finetuned Petals, Bloomz, BERT RAG LLMs and optimized LLM inference using vLLM

Flock - Inter Vehicular Communication with Deep Learning, IoT (Ctrl+Alt+Code National level Hackathon Winner)

- Assembled a prototype to control multi-vehicles by a single driver by platooning using GPS and Ultrasonic sensors
- Finetuned a ResNet CNN model to detect objects around the vehicle | Github | Demo video