**Education**

**Masters of Computer Science Sept 2022 – June 2024 (Expected)**

University of California San Diego (UCSD) CGPA: 3.97/ 4

*Relevant Courses*: Software Engineering, Advanced Data-Driven NLP, Deep Learning, Scalable Data/ML Systems, Recommender Systems, Bioinformatics / Genetics

**Bachelor Of Engineering (Computer Engineering) August 2018 – July 2022**

Vivekanand Education Society’s Institute of Technology (VESIT) CGPA: 9.013/ 10

*Relevant Courses*: Natural Language Processing, Database Management Systems, Web Development, Operating Systems, Machine Learning, Software Engineering, Big Data (Hadoop) Analytics, Cloud Computing, Object Oriented Programming (Java)

**Internship Experience**

*Machine Learning NLP and CV,* **Ujima S&P Lab UCSD**  **March 2023 – Present**

* Building the AWS Cloud Infrastructure to host a Flask based LLM User Study on AWS Beanstalk with a Mongo DB backend.
* Building a “Smart Mirror” on Raspberry Pi using VGG-Face TF-Lite Model, trained on 100,000 images, to detect ethnicity, address the existing biases in CV datasets, and contribute to the public domain via user feedback.
* Developed a rule-based security chat assistant using Rasa in an unstructured, fast-paced startup-like environment.
* Led 4 undergraduates through the Early Research Student Program in analyzing privacy data for MMO AR/VR games using TF-IDF and K-means Clustering.
* Led 3 freshers through the Early Research Program in curating a database using LLMs, Web scraping and NER for HCI Research.

*Data Analyst & Cyber Security Intern,* **Legendary Entertainment**  **June 2023 – Dec 2023**

* Created a Python automation script to generate dashboards that reduced the time taken by the SOC analyst from 80 minutes to about only 4 minutes daily in classifying user-anomaly events by integrating Splunk, Active Directory and Azure MSGraphs APIs.
* Maintained a high level of security while streamlining a user’s authentication process by integrating a FIDO Alliance product into the existing SAML/OIDC SSO workflow.
* Assisted the VP of security in laying the foundation of a next-gen universal anomaly and user-behavior detection platform based on LLMs in collaboration with Sky High Security by building a prototype in Splunk.

*Undergraduate Research Assistant,* **Tata Institute of Fundamental Research (TIFR)**  **June 2021 – May 2022**

* Led a team of 4 to develop an android Java application that monitors a selected directory and uses multi-part upload methodologies to encrypt and securely upload to the dedicated remote server.
* Published a [paper](https://doi.org/10.1007/978-3-031-18497-0_41) explaining our Node JS based Fault Tolerant client server architecture connected to remote stations.
* Utilized GCP’s Maps and Sheets API to design a real time HTML/CSS based live Geo tracking website from the collected data.

*Full Stack Development Intern,* **Makos Infotech (Startup)**  **June 2021 – August 2021**

* Integrated and developed server-side code using JQuery, PHP, and MySQL for an early startup targeting the automation of the On-campus placement process ~~using Scrum framework/agile methodologies~~.
* Created and merged relational databases using MySQL workbench and deployed it on AWS RDS to develop a college-student-company social network inspired by Facebook’s friend system.
* Established a mentorship-onboarding program for new undergraduate interns, aligning them with the existing codebase and processes, saving the company at least 1 week of time and effort.
* Pitched and demoed the online job placement portal to 3 University board members alongside the founding CEO.

*Web Developer Intern*, **VESIT Renaissance Cell** **June 2020 – July 2020**

* Led the design and development of a Django-based [Paper Publication Web service](https://github.com/JayJhaveri1906/Django_Website_publications/), implementing essential CRUD operations, for 50 professors hosted on [Heroku](https://cmpn-publications-official.herokuapp.com/), serving approximately 250 CSE students.
* Developed a [Portfolio Website template](https://jayjhaveri190600.web.app/) using HTML/CSS/JS and Bootstrap, deployed on Google’s Firebase.

*Data Analyst Intern*, **Leadingindia.ai** **May 2020 – July 2020**

* Collaborated with a team of four to develop a vaccine prediction model for H1N1 and seasonal flu vaccines, accurately predicting public acceptance trends (41%) for the COVID-19 vaccine, securing first place among 85 intercollege peer groups.
* Published a [research paper](https://doi.org/10.1007/978-981-16-0401-0_11) in Springer & authored a [blog](https://medium.com/@jjhaveri1906/pandemics-a-harsh-reality-7c05254e907b) highlighting the correlation between H1N1 and COVID-19 pandemics.

**Projects**

**[Global Ancestry Comparison (Bioinformatics)](https://github.com/NirajYagnik/Global-Ancestry-Comparison/?tab=readme-ov-file" \l "global-ancestry-comparison)** [(Presentation)](https://www.canva.com/design/DAF_VmEn2ck/VWlvkANMqdH2hzRMXhvdFw/edit) **Jan 2024 – March 2024**

* Led the development and benchmarking of multiple dimensionality reduction techniques such as PCA, t-SNE, UMAP, MDS, Isomap, and Autoencoders for global ancestry inference on 1000 Genome project’s VCF files.
* Awarded the best class presentation for in-depth analysis by using homogeneity scores to compare the techniques.

[**Inquirable Models: Increasing Explainability in Health-AI using LLM**](https://docs.google.com/presentation/d/1naydNzz6F8W51bA40Phez4-Pj2b-vJRJmUHkOp5rO1M/edit?usp=sharing) **Sep 2023 – Jan 2024**

* Conducted a two-phase exploratory study using prompt engineering techniques on leading Large Language Models (LLMs) with SHAP values to improve the interpretability of traditional medical risk models and reduce patient risk.
* Facilitated surveys with doctors and patients to assess answer quality, focusing on metrics such as confabulation rate.
* Paper’s poster accepted for presentation at the AMIA 2024 Annual Symposium.

[**Alt Bot for Mastadon: An automatic image alt generation bot**](https://github.com/CSE210-Fall23-Team2/AltBot)[(Presentation)](https://docs.google.com/presentation/d/1smZzOd8u-NhgbotJRkn2Eqw5WmXVxA-XSB0afzxaNWE/edit?usp=sharing) **Sep 2023 – Dec 2023**

* Developed a REST-API based Chrome extension in JS to help visually impaired people browse decentralized social media feeds by leveraging hugging face image captioning models to generate alternative image descriptions.
* Deployed 3 levels of custom cache system to ensure efficient performance with minimal lag complemented with testing scripts.

[**MedLM: Exploring Language Models for Medical QnA Systems**](https://github.com/JayJhaveri1906/CSE291_MedLM)[(Paper)](https://arxiv.org/abs/2401.11389) **March 2023 – Aug 2023**

* Led a team of 4 to fine-tune language models (Bloom, T5, GPT-2) on the MedQuad dataset in collaboration with Microsoft researcher Dr. Asma Ben Abacha.
* Compared performance against GPT-3.5 and GPT-4 using Dynamic Prompting with Retrieval Augmented Generation (RAG) via medical InstructOR Embeddings on the patient questions.
* Increased the ROUGE and BLEU scores by 10% by using a Bert Classifier to give extra contextual awareness the models.

**[GrooveGenie: A copyright-free music generator](https://github.com/JayJhaveri1906/GrooveGenie-A-copyright-free-music-generator) March 2023 – June 2023**

* Created an open-source music generation model, utilizing Facebook’s EnCodec Transformer model to compress audio wav files into embeddings for model interpretation.
* Trained a conditioned Multimodal GAN network to generate music based on user-provided genre inputs embedded using the BERT model, aimed at producing only copyright and royalty-free music, trained on the FMA dataset.

**[Game Genre and Recommendation Classification using Steam Reviews](https://github.com/JayJhaveri1906/Game-Genre-and-Recommendation-Prediction) Nov 2022 – Dec 2022**

* Designed data pipelines to preprocess and apply machine learning techniques for classifying game genres, analyzing user sentiment, and curating a personalized game recommendation system using user reviews.
* Achieved 90.53% accuracy with Random Forest, balanced data & TF-IDF, outperforming N-Gram, Multinomial NB, Linear SVC.

**[Aatmanirbhar Sanchar: Secure Self-Sufficient Communications](https://github.com/JayJhaveri1906/Aatmanirbhar-Sanchar)** [(Paper)](https://doi.org/10.1007/978-3-031-18497-0_41) **June 2021 – May 2022**

* Led a team of 4 to develop a secure, end-to-end encrypted, multimedia-supported chat application for the Tata Institute of Fundamental Research Institute (TIFR), integrating off-the-grid functionality for internal use.
* Utilized CI/CD practices to establish a client-server architecture leveraging Python and Node, resulting in streamlined SDLC.
* Collaborated in implementing SHA-256 and AES-256 overlapped inside an HMAC envelope to protect against cyber attacks.

[**Aatmanirbhar Sanchar: Secure Self-Sufficient Communications**](https://github.com/JayJhaveri1906/Aatmanirbhar-Sanchar)[(Paper)](https://doi.org/10.1007/978-3-031-18497-0_41) **June 2021 – May 2022**

* Led a team of 4 to develop a secure, end-to-end encrypted chat application with off-the-grid functionality for the Tata Institute of Fundamental Research Institute (TIFR), utilizing CI/CD practices to establish a streamlined client-server architecture.
* Collaborated in implementing SHA-256 and AES-256 overlapped inside an HMAC envelope to protect against cyber attacks.

[**Divya-Drishti: An Independent Aid for the Visually Impaired**](https://github.com/JayJhaveri1906/Divya-Drishti)[(Paper)](https://dx.doi.org/10.2139/ssrn.3867707) **Aug 2020 – May 2021**

* Achieved a 400% net cost reduction by creating a real-time Voice-activated AI IoT android application to help Visually Impaired People (VIPs) comparable to state-of-the-art OrCam in multimodal classification of currency, objects utilizing GCP’s Vertex AI.
* Published a [research paper](https://dx.doi.org/10.2139/ssrn.3867707) highlighting the needs of VIPs, funded by the Mumbai University Minor Research Grant.

[**Automated Number Plate Recognition and Parking System**](https://github.com/JayJhaveri1906/AutomaticParkingSystemANPR) **Dec 2019 – Feb 2020**

* Built an Android Java based application connected to a Firebase server to automate security and space availability in car parking systems by monitoring the number plates detected at the exits using Tesseract OCR.
* Dockerized the software to streamline deployment on existing CCTVs at parking lot gates, minimizing upfront costs.

[**International Flutter Hackathon: Healthy While Distant**](https://github.com/JayJhaveri1906/Healthy-While-Distant) **June 2020 - 48 hours**

* Devised a user-friendly Flutter app that leveraged smartphones' existing Bluetooth Low Energy (BLE) technology to help users maintain social distancing during the COVID-19 pandemic and alerted on a violation.
* It was combined with an additional feature of teaching yoga positions to promote positive mental health.

**Selected Research Publications**

Inampudi S., **Jhaveri J.** et al., (2021) **Machine Learning Based Prediction of H1N1 and Seasonal Flu Vaccination**. Advanced Computing. IACC 2020. Communications in CIS, vol 1367. Springer, Singapore. [DOI.org Link](https://doi.org/10.1007/978-981-16-0401-0_11)

- *AWS Certified Cloud Practitioner (CCP)*, Amazon Web Services, [Credly Link](https://www.credly.com/badges/45871348-ade7-4b07-89ee-6eeb7e85b72f/linked_in_profile)

- *Machine Learning Based Prediction of H1N1 and Seasonal Flu Vaccination*. Advanced Computing. IACC 2020. Springer. [DOI.org Link](https://doi.org/10.1007/978-981-16-0401-0_11)

**Skills:** Python, SQL, Java, Javascript, HTML/CSS, React, Django, PostgreSQL, Android, Linux, Git, AWS, Azure, GCP, Firebase, Hugging Face, Docker, PyTorch, Tensorflow, OpenCV2, Scikit-Learn, Pandas, Splunk, SPL, Object Oriented Programming, APIs, Android-Studio, Android-Java, Flutter, C,