Aditya Vadalkar

vadalkar@usc.edu | +1 (213) 536-3281 | [LinkedIn](#) | [Github](#) | [adityavadalkar.github.io](#)

EDUCATION

MS in Computer Science	University of Southern California	Aug 22 - May 24
Focus: Web Development/ML/NLP	**Courses:** Web Technologies, Algorithms, Predictive Analytics	
BE in Computer Engineering	University of Pune, India	Jul 18 - Jul 22

SKILLS

Programming Languages: Python, Javascript, HTML, CSS (Proficient); Java, C++, C#, TypeScript (Prior Knowledge)

Web Technologies: React, Vue, Node, Flask, Shiny, Django, Redux, Figma, Material UI, Web.gl, Three.js

Databases: Postgres, MySQL, SQL, MongoDB, Pinecone, ChromaDB

Techniques: A/B Testing, Agile, UI/UX Principles, Software Design Patterns

Machine Learning: Transformers, QLoRA, LLM, RNN, DNN, LSTM, Decision Trees, KMeans, Random Forest, XGBoost, SVM

Libraries: Pytorch, Tensorflow, PEFT, TRL, Numpy, Pandas, Matplotlib, SKlearn, Langchain, OpenCV, Seaborn, OpenAI

Cloud & Version Control: AWS, Git, Google Cloud, Heroku

Tools: Docker, Trello, Notion, Docker, Unity

EXPERIENCE

RTI International - Los Angeles, CA | Data Scientist (previously AI/Data Science Intern)   Jul 23 - Current
- Implemented regression models (CatBoost, XGBoost, Random Forest) to explain 99.5% variance in opioid mortality in US.
- Created visualizations (SHAP plots, partial dependence plots, scatter plots) for increased opioid model interpretability.
- Used FLIR SDK and Exiftool for thermal metadata extraction from FLIR thermal videos.
- Implemented optimized TF Lite models for CPU-based cognitive load detection on facial ROI.
- Developed interactive dashboards for COVID-19 and Drug Overdose with R Shiny to analyze social determinants of health.
- Optimized the dashboards with asynchronous loading, data chunking, and MapShaper tool; deployed to ShinyApps.io.

Raven: Fashion AI - Los Angeles, CA | Founding Software Engineer GenAI ([Demo](#))   Jan 23 - Oct 23
- Engineered a custom AI conversational agent using Streamlit, ChatGPT API, Langchain, and Pinecone vector database.
- Created custom agents and tools in Langchain and created a querying mechanism using Pinecone's similarity search.
- Populated vector database with indexes for ~1 million products using FAISS and Huggingface SentenceBert.
- Designed and developed the entire front end for Raven using React, Redux, and Figma.
- Designed backend Node.js APIs and deployed website to Google Cloud Platform through collaboration.
- Increased user acceptance ratings by 15% through several UI/UX changes & leveraging over 200 customer insights.
- Implemented secure user authentication using OAuth2.0 and engineered unit tests using Jest.

Athabasca University - Remote | NLP Research Intern   Jun 21 - Aug 22
- Augmented Hierarchical Attention Networks boosting citation detection/localization accuracy/F1 by 4% on Open Corpus.
- Demonstrated Gated Multilayer Perceptrons' effectiveness in text alignment on ACL Anthology Network/S2ORC datasets.
- Performed qualitative analysis using attention visualization to identify their strengths and weaknesses in text alignment.

Tech Mahindra - Pune, India | Machine Learning Intern   Jun 20 - Sep 20
- Developed end-to-end pipeline for automated text-to-2D animation conversion.
- Utilized Sentence-BERT for text encoding, and facilitated training of GANs using these sentence vector representations.
- Curated new labeled datasets for 2D and 3D videos to enhance text-to-video conversion performance.

PROJECTS

AI-Powered Sports Highlights Generator ([Demo](#))
- Led a team of 5 in developing a Retrieval-Augmented Generation (RAG) system with LangChain and OpenAI GPT-4 models for generating relevant highlights.
- Secured 3rd place in multimodal hackathon hosted by Twelve Labs with judges from AWS, NFL, Twelve Labs, and FBRC.ai.
- Devised a pipeline to download and transcribe sports press conference videos using PyTube and AWS Transcribe.
- Utilized AWS Rekognition for speaker identification and attached speaker metadata to the transcript for diarization.
- Chunked and indexed the transcripts for efficient retrieval, storing data in a vector database using SQLite.
- Built a web application using Python and Flask, providing an intuitive interface for users to interact with the tool.

Step-by-Step Distillation - Llama to DialoGPT
- Utilized the QLoRA technique to finetune Llama7B on the *Empathetic Dialogues* dataset from Huggingface.
- Leveraged the fine-tuned model to generate rationales and curated a new empathetic dialogues (with rationales) dataset.
- Fine-tuned DialoGPT (147M) on curated dataset via step-by-step distillation, resulting in a 30% improvement.

Intelligent Tutoring System
- Developed an NLP-driven e-learning platform for content summarization, question generation, and keyword extraction.
- Implemented TextRank algorithm for extractive text summarization of reference text into important notes.
- Leveraged T5 model for automatic multiple-choice question and answer generation from reference texts. Achieved an average score of 4.2 (out of 5) on diverse NLG evaluation criteria.
- Integrated YAKE to highlight keywords, hyperlinked to external sources for further reading for the user.
- Deployed the pipeline to Google Cloud Platform and developed REST API endpoints using Python and Flask.

Lightweight Yelp
- Developed a website for optimized business detail retrieval with React by utilizing the Yelp API.
- Created a backend using Node.js and Express framework with 5 REST API endpoints.
- Collected location data using Google Geocoding API. Displayed maps and Yelp reviews using Google Maps and Yelp API.
- Engineered a similar Android app using Java, Volley, & Picasso, improving user experience and increased app efficiency.

Twitter Sentiment Browser Extension
- Engineered a Node.js server to generate sentiment scores using Huggingface models for tweets on the Twitter website.
- Manipulated the DOM using Javascript, HTML, & CSS to display emojis next to tweets depending on sentiment scores.

Asthma and Diabetes Predictor
- Developed ML pipeline using a suite of regression models achieving R2_score ~0.95 for diabetes and asthma prediction.
- Designed user-friendly interface in kepler.gl for intuitive navigation and user understanding.
- Presented at AI LA Open Life Summit 2022, won AILA Open Innovation Challenge 2022.

Covid Spy
- Devised a prototype for contact tracing by employing Bluetooth Low Energy Beacons (BLE) and RSSI.
- Implemented graph data structure to stratify afflicted people into social networks
- Ideated a risk infection hierarchy structure to allow for easy tracking of potentially infected individuals

ACHIEVEMENTS

- Winner of the Twelve Labs: Multimodal AI Media & Entertainment Hack organized by Twelve Labs, FBRC.ai and AI LA, securing 3rd place for our solution AI-Powered Sports Highlights Generator.
- WInner of the COVID-19 Global Hackathon 1.0 for our solution of Covid Tracker