Anay Modi

4540 Oakmont Ct., Shelby Township, MI 48317 | 586-850-4437 | anayvmodi@gmail.com
linkedin.com/in/anay-modi | github.com/anay-m | anay-m.github.io

EDUCATION

University of Michigan – College of Engineering, Ann Arbor, MI May 2026
B.S.E. in Computer Science + Minor in Mathematics

- **GPA**: 4.00/4.00 | **Coursework**: Data Structures & Algorithms, XR Development (VR & AR), Operating Systems, Programming Paradigms, Machine Learning, Computer Organization, Linear Algebra, Probability Theory
- **Relevant Activities**: Michigan Data Science Team, V1 Michigan, Atlas Consulting, Alternate Reality Initiative
- **Interests**: Avid NBA, NFL, & NCAA Fan, Compete in Flag Football league, Former Bollywood Dancer

SKILLS

Languages: Python, C++, SQL, JavaScript, HTML/CSS, TypeScript, C#, Java, Swift, Golang
Frameworks: Unity XR, React.js, Node.js, PyTorch, TensorFlow, Pandas, LangChain, Unreal Engine, CUDA, Angular.js
Technologies: Git, Linux, AWS, Google Cloud, Azure, Spark, Kafka, Kubernetes, Redis, Pinecone, MySQL, Jira
Certifications: J.P. Morgan Chase Virtual SWE Exp., Machine Learning (Stanford), XR & Web Development (UMich)

EXPERIENCE

Epic Systems: *Software Development Intern,* Madison, WI May 2024 – Present

- Architected a **time-series** machine learning model using **PyTorch & Scikit-learn** to forecast # of unfilled appointment slots in providers' schedules, recording **90%** accuracy on training data for over **100,000** providers.
- Researched and validated various forecasting models, including regression, tree-based models, gradient boosting, Markov Models, and Recurrent Neural Networks, utilizing Optuna and grid search for hyperparameter tuning.
- Deployed and trained models on **Azure-based cloud computing** platform, integrated with a **TypeScript** frontend & distributed event streaming pipeline via **SQL** to introduce **real-time** schedule forecasting on scheduling app.

Diagnostic Intelligence Augmentation for Global Health: *ML Researcher,* Ann Arbor, MI Jan 2023 – April 2024

- Supervised 5-person developer team at Dr. Arvind Rao's AI-cancer research lab, utilizing **ImageNet CNN** models & **transfer learning** to identify tumors in Whole Slide Images (WSI's). Managed large-scale data storage using **Google Cloud** buckets and leveraged **GPU** clusters for model training, achieving **95+%** accuracy in simulations.
- Conducted analysis and documentation of **Python** deep learning libraries, spearheading the development of a WSI Deep Learning **Docker** Package for health organizations looking to deploy custom models for tumor prognosis.

U-M Information & Technology Services: *Software Engineering Intern,* Ann Arbor, MI May 2023 – Aug 2023

- Launched migration of UM API middleware software from WSO2 to Denodo, engineering novel API prototypes within Denodo **REST** service and Google Apigee console and reducing API management costs by **$30K.**
- Designed **SQL** views to expose data from UM databases as APIs, deploying backend **JavaScript** programs to connect UM APIs to Denodo server while enabling memory caching & pagination for over **2x performance.**

PROJECTS

Nostalgia: *Full Stack Developer / AI Engineer [React Native, AWS, Kubernetes, Redis, Pinecone, PyTorch, Figma]*

- Developed a **React Native iOS/Android** app that allows users to search and share memories across social media apps, including Apple Photos, Google Photos, Instagram, Snapchat, VSCO, TikTok, Netflix, YouTube, and Spotify.
- Architected a scalable backend infrastructure using **AWS** services (Lambda, EC2, SQS, DynamoDB, S3, CloudWatch), **Kubernetes**, and **Redis** to support over **50** test users, with plans for further expansion.
- Integrated **FaceNet** and **CLIP** models to accurately identify and label people, objects, and landmarks in multimedia content. Developed a semantic search engine and chatbot using **OpenAI, Pinecone** and **LangChain**, enabling users to explore & share memories through natural language (i.e. "show the last time my brother & I played soccer").

WallStreetBets NLP Analysis: *Full Stack Dev / AI Engineer [React.Js, TensorFlow, AWS, Firebase, Pinecone]*

- Engineered a **React.js** web app allowing users to search for stocks and view relevant financial data, Reddit posts from WallStreetBets subreddit, and percentage of returned posts with positive, negative, and neutral sentiments.
- Implemented Python scripts in **AWS** to classify posts from WallStreetBets according to their sentiment using **BERT NLP analysis**, updating **Pinecone** databases so users can view relevant Reddit posts via semantic search.

C++ Thread Library: *Software Developer [C++, Linux, Unix]*

- Authored a **C++ 17** thread library on **Linux**, handling CPU initialization, thread creation for 50+ CPUs, interrupt handling, and atomic operations, along with scheduling and synchronized context processes.

Sora Detection Analysis: *Machine Learning Engineer [PyTorch, Optuna, Jupyter, Google Cloud]*

- Engineered ensemble of **convolutional neural networks** to recognize AI-generated videos from Sora, Make-A-Video, and the Deepfake Detection Challenge, using **ELA** preprocessing & **PyTorch** for over **85%** test accuracy.