HARSHINI SAI DONEPUDI

New York, NY | 702-820-9381 | US Citizen
github.com/HarshiniDonepudi | Linkedin | hsd39@cornell.edu

EDUCATION

Cornell Tech (Cornell University), New York, NY |Merit Scholar| GPA: 4.0 May 2025
Jacobs Technion – Cornell Dual Master of Science Applied Information Science and Information systems – Health Tech Concentration
Computer Vision, Machine Learning Engineering, Data Science in the Wild, Deep Learning, User Interface Design, Web Developement

Andhra University College of Engineering (A), Visakhapatnam, India | Best Outgoing Student '23|GPA: 3.92 May 2023
Bachelor of Science in Computer Science and Systems Engineering
Machine Learning/AI, Cloud Computing, Data Structures and Algos, Object Oriented Programming, Software Engineering, Operating Systems

TECHNICAL SKILLS

Coding Language: Python, R, SQL, Kotlin, Swift, C++,Javascript, MATLAB, Objective-C, Java, HTML, CSS, SAS
Frameworks: Tensorflow, Pytorch, LangChain, HuggingFace, Pandas, Scikit-Learn, Numpy, Matplotlib, Pyspark, Streamlit, Flask
Tools: Xcode, Android Studio, VScode, Git, Azure, AWS, GCP, Databricks, Snowflake, Tableau, PowerBI, Figma, MS Office

EXPERIENCE

Quest Diagnostics – Data Scientist | Remote Sept 2024 – Present
- Developed predictive models using Python, PySpark, SAS and SQL for risk stratification and clinical insights and built ETL pipelines for large-scale lab and medical coding datasets (ICD-10 and CPT) using Snowflake and optimized data transformations.
- Collaborating with engineering teams to build scalable AI/ML pipelines in cloud environments (AWS, GCP), utilizing NLP and deep learning to improve operations and created data visualizations with AWS QuickSight and Snowpark for efficient data management.

Weill Cornell Medicine – Machine Learning Researcher, New York, NY Sept 2023 – Present
- Developed and deployed AI-driven medical imaging solutions (PyTorch: UNet) to improve heart donor-recipient matching and integrated LLM's and NLP techniques (NER) for clinical text retrieval and document intelligence improving accuracy by 10%.
- Employed experiment design and statistical measurement (A/B tests, causal inference) for real-time DICOM to NifTi preprocessing pipelines on HPC clusters and Azure, effectively communicating complex technical findings to clinical teams to support decisions.

GE Healthcare – EEDP Software Intern, New York, NY May 2024- August 2024
- Optimized cloud resource usage for GE Healthcare's Portrait mobile using FinOps/DevOps techniques. Automated trend analysis dashboards with the TICK stack, Python, and SQL, integrated into CI/CD pipelines via Jenkins for seamless 12-hour updates.
- Collaborated in cross-functional stakeholder teams to align technical insights with business objectives enhancing efficiency by 15%.

Johnson and Johnson – Innovation Medicine – Commercial Data Science Coop, Titusville, NJ Jan 2024 – May 2024
- Developed a multimodal GenAI-powered LLM with RAG (Hugging Face, LangChain) to automate insight extraction from healthcare claims, insurance, pharmacy and drug SHAP plots, to generate data stories—supported by prompt engineering.
- Employed experiment design and A/B testing techniques to drive strategic recommendations and enhance HCP Salesforce outputs by 25%; built MLOps data pipelines with AWS S3, SageMaker, and PACS (HL7/FHIR) for robust processing.

Mayo Clinic, Intern – Biostatistics and Machine Learning, Rochester, MN March 2023 – Jan 2024
- Designed AI pipelines combining CNNs and Vision Transformers (VIT)(Pytorch, Tensorflow) to analyze multimodal dataset (imaging and tabular), data processing using R and Python, collaborating with stakeholders for real-world validation.
- Applied GPU Optimisation, fine tuning and parallel processing for an accuracy improvement of 4% on HPC Clusters.

Shibnobi – Software, Technology Developer, Azle, TX August 2022 – Jan 2024
- Developed cross platform applications for projects: Shibnobi Play, Foundation Token, Torqued and Tuvoznow for IOS and Android.
- Led the end-to-end development of AI-driven cross-platform applications, implementing gaming-based predictive analytics, Tableau dashboards for user insights and NLP models to enhance user engagement and content recommendations.

PROJECTS

- **Wound Whisperer** *(*Freelance Project for Wound Expert Care Ltd | Python, Azure Databricks, PyTorch, SQL*).* July 2024 – Ongoing
  Developed an AI multi-modal tool for wound detection and severity assessment using Azure Databricks. Built a custom wound annotation pipeline with Node.js and Claude. Processed 200,000+ images from EHR data for segmentation and classification. Integrated CoreML for real-time iOS deployment, reducing latency by 25%, and added body mapping, and healing progress visualization.
- **InsuloCarb** (Python, CoreML, GenAI, TensorRT, Swift, HealthKit, ARKit). April 2022 – Ongoing
  Developed a computer vision app using ARKit for volume estimation and AI to estimate meal carbohydrates and recommend insulin doses. Enhanced glucose predictions 2-4 hours in advance, factoring in meal sensitivity for precise dosing. Integrated GenAI (GPT-4o) for tailored meal suggestions and synced with Dexcom systems and Insulin Pumps for real-time compatibility and alerts.
- **Clinical Fact Verification System using RAV's** (Python, LangChain, RAG, OpenAI API, Streamlit, Pytorch) Aug 2024 - Dec 2024
  AI system for verifying medical claims and generating evidence-backed claim using RAG and LangChain. Integrated GPT-4o with Streamlit to automate claim validation against trusted medical sources in vector db, improving accuracy and decision-making efficiency

PUBLICATIONS

- A Comprehensive Study on Accident Detection Techniques - DELCON '22(IEEE conference), IEEE Xplore April 2022
- Insulin Usage and Practices in Children and Adolescents with Type 1 Diabetes Endocrinology and Metabolism July 2021