

SAIKIRAN PUVVADA

Portfolio

puvvadasaikiran

saikirann.puvvada@gmail.com

puvvadasaikiran

+91-7658970090

SUMMARY

Data Science professional with 5 years of proven experience delivering successful end-to-end NLP, Voice, and Speech AI solutions in the US Healthcare domain. Adept at understanding user needs and motivations to design intuitive solutions, effectively addressing complex challenges through empathetic communication.

Currently Looking for opportunities as a Senior Data Scientist / Consultant / AI Product Manager

SKILLS

AI Modelling: Pytorch, Tensorflow, Nemo, Kaldi, EspNet, SpeechBrain, HF Transformers, LangChain, RAG, LLM Finetuning

Programming/ Scripting Skills: Python, JAVA, Bash, ReactJS, VueJS, NodeJS

Statistical Analysis: Descriptive Statistics, Hypothesis Testing

Database: SQL/MYSQL, MongoDB, DuckDB

Deployments: Kafka, FastAPI, Flask, SocketIO, AMPQ, Pub/Sub, Microservices

Data Scraping and Analysis: Pandas, NumPy, BeautifulSoup, Selenium, ScraPy

Data Visualizations: Tableau, Excel, Matplotlib, Power BI, Illustrator

Platforms: Microsoft Office Suite, GCP, AWS, Microsoft Azure

Version Controlling Tools: Git/GitHub, BitBucket

Product Management Tools: Monday.com, Jira, Confluence, Asana

Data Engineering Tools: Docker, Kubernetes, Kafka, PySpark, Pub-Sub, AMPQ 0-9-1, Celery, MLFlow, TensorBoard

UI/UX - Rapid Prototyping: Figma, Adobe XD, Gradio, Streamlit

EXPERIENCE

Lead Data Scientist (Optimization) at Genzeon [Sept 2025 : Present]

- Spearheaded adoption of **spec-driven development with OpenSpec** across the platform, cutting feature delivery timelines by **70%** through earlier alignment between engineering, product, and clinical stakeholders
- Established **cloud-agnostic, CI/CD-driven deployment standards** with per-client infrastructure isolation, meeting **HIPAA and SOC 2** compliance requirements while preserving portability across cloud providers
- **Client Onboarding Acceleration : In-Patient Precision Care Platform**
 - Cut client onboarding time from **45 days to 10 days** by designing a rules-based mapping engine that translates heterogeneous client data into the platform's internal representation, paired with on-demand provisioning of **isolated per-client infrastructure** for compliance-driven tenancy
 - Built containerized migration workflows enabling clinics and health centers to be provisioned independently without blocking on shared infrastructure
- **Resilient Data & AI Processing Engine (HL7 & FHIR)**
 - Architected an **event-driven, Temporal-orchestrated** processing engine transforming raw HL7 and FHIR feeds into unified internal representations, replacing poll-based workflows with reactive pipelines and built-in recovery, observability, and governance controls
 - Refactored legacy monolith into microservices using the **SAGA pattern** over Temporal, enabling fault-tolerant distributed transactions across ingestion, enrichment, and AI inference stages
 - Reduced pipeline compute footprint by **50%** by migrating workloads from always-on VM-based processing to **containerized AKS deployments with KEDA-driven autoscaling**
 - Reduced LLM inference costs through prompt caching, batched bulk operations, and lazy evaluation to eliminate redundant API calls across enrichment workflows
 - Decreased clinician-reported incidents on AI recommendations by **40%** by improving Medical AI Agent factuality through hallucination-reduction techniques and RAG-based grounding against curated clinical knowledge sources
 - Designed tiered archival policies aligned to clinical data retention requirements, balancing query performance against long-term storage costs
 - Engineered alignment strategies between HL7 v2 and FHIR data models, resolving schema and semantic mismatches to ensure consistent downstream representation

Senior Data Scientist (Speech/NLP) at Edvak Health [Apr 2023 : Sept 2025]

• Real Time Conversation Capture Stack for US Healthcare

- Designed and Implemented scalable Conversation Capture Stack for EHR to reduce documentation burden on providers by 50% (Python, SocketIO, sklearn, Pytorch)
- Led the deployment of the system from scratch, using Websockets to ensure scalability and reliability (Kubernetes, CI/CD, Docker)
- Guided a 4-member team of junior data scientists in the modeling process, using advanced speech processing techniques and Large language Models(LLM) (SpeechBrain, PyAnnote, Clustering, Role identification, GPT4, DeepSeek, Llama, OpenAI)
- Successfully Integrated NLP/text dependent role identification techniques to improve on relative speaker confusion rate by 28% in a conversation

• Front Desk call Automation for US Healthcare

- Designed and Implemented Call Automation for US Healthcare with VITS based Text to Speech Synthesis and LLM for appropriate response generation.
- Designed scalable deployment strategy to handle patient queries over call in place of front desk staff.
- Utilized LLM's Natural Language Understanding and Natural Language generation capabilities for response generation using prompting techniques.
- Utilized Large Action Models to perform actions based on patient's conversation flow appropriately.
- Working with Livekit, Twilio for seamless scalable integration into clinic's existing phone systems

• Auto-Punctuation and Capitalization Model for US Healthcare

- Developed BERT based Auto-Punctuation and Capitalization Model for improving clinical document readability
- Utilized Data Augmentation/Generation/Scraping techniques to tackle data scarcity (Pandas, Regex, Selenium)
- Improved punctuations and capitalization performance by 24% over base model using re-defined modelling objective for in-domain transfer learning.

• Edvak EHR to HL7 FHIR Data Mapping

- Spearheaded the Healthcare Data Interoperability Initiatives in the organization to transform internal data representations into HL7 FHIR standard for improved Population Health Management, Care Coordination and CDSS Services. (Python, MongoDB, HL7 FHIR, CDSHooks)

• Managerial Responsibilities

- Co-ordinating with the AI and Integration Teams to optimize the work plan to align towards long-term Organizational Business Objectives.
- Managing and guiding junior data scientists throughout the end-to-end project life cycle.
- Communicate with stakeholders to discuss the plan / progress.
- Managing and planning annotation activities for speech / NLP use cases
- Exploring transformative AI initiatives tailored to emerging use cases in healthcare, with an emphasis on product discovery and future-forward solutions.

Data Scientist (Speech/NLP) at Edvak Health [Dec 2021 : Apr 2023]

• Automatic Speech Recognition(ASR) for US Healthcare

- Worked on developing Closed Language Model Automatic Speech Recognition Engine for US Healthcare.
- Involves working with Machine Learning Libraries PyTorch and TensorFlow in Python and Kaldi framework for Automatic Speech Recognition (Kaldi, Wav2Vec2, RNN-T, Parakeet, Whisper).
- Setting up near-realtime ASR Inference endpoints at large scale (Socket-io, Kafka, Websockets, RabbitMQ)
- Hyper Parameter tuning the models for better performance (Machine Learning, Deep Neural Networks).
- Handling Raw Audio Data and Processing it to eliminate abnormal data (Sox, Bash, Python, Data Analysis, Pandas, Numpy, Librosa).
- Developing In-Domain Benchmarking Test-set for Rapid Model Evaluation.
- Automating Data Collection for Active Learning and Strategizing Data Strategy at Organisation Level.

• Language Modelling [n-gram] for ASR contextual biasing

- Cleaning and Pre-processing Text Data at large scale (>10GB) for N-gram Language Model training using SRILM/KenLM toolkit(Pandas, Regex, Bash, Python, GNU-Parallel, Dask, OpenRefine).
- Designing/Developing/Managing Pipelines with lexicon filtering and Web scraping for Automated LM Generation and Evaluation (Perplexity, Bash, Cron jobs, Selenium, ScraPy).
- Managing metadata Databases for Data Versioning and Data Lineage (SQL, MongoDB)

EDUCATION

Indian Institute of Technology Madras

October 2020 - December 2023

B. S in Data Science and Applications

Sreenidhi Institute of Science and Technology, Hyderabad

July 2016 - June 2018 (Matriculated)

B.Tech in Computer Science and Engineering

CERTIFICATIONS

- **AI in US Healthcare Specialization by Stanford University** | X18B7H5NI7B4 | Coursera 
- **Deep Learning Specialization** | T9P329J56AJ8 | Coursera 
- **Machine Learning Engineering for Production (MLOps) Specialization** | 3WEC7TGKXHJN | Coursera 
- **Natural Language Processing Specialization** | JWBBB93TSK7Q | Coursera 

HONORS / ACHIEVEMENTS / VOLUNTEERING / MENTIONS

- **Lead Instructor | Machine Learning Operations (MLOps)**  | IIT-Madras (Mar 2025 - Sept 2025)
 - **Concepts** : Git, DVC, Feature Stores, CI/CD, K8s, Docker, Data Dog, Prometheus, Grafana, MLFlow, SHAP
- **Teaching Assistant | Introduction to Big Data**  | IIT-Madras (Jan 2024 - Apr 2025)
 - **Concepts** : Spark, Kafka, Serverless Deployments, Pub/Sub, Spark MLlib, GCP Cloud Essentials
- **Experience as a freelance flutter and UI/UX Developer (2018-2020)**
- **Technical Head (campus role at ArtsClub, SNIST) (2017-2018)**
- **Volunteered for Global Entrepreneurship Summit, Hyderabad, 2018**
- **Indian Government Undergraduate National Scholarship Awardee (2016-2018)**