

Prabhakar Elavala

San Francisco, CA · +1 8573980726 · [Gmail](#) · [LinkedIn](#) · [GitHub](#) · [Portfolio](#)

Education

M.S - Informatics (Concentration: Analytics) Northeastern University, Boston, MA	January 2024 – December 2025 GPA: 3.8
B. Tech - Electronics and Communication Engineering Sree Nidhi Institute of Science and Technology (SNIST), Hyderabad, India	August 2019 – June 2023

Work Experience

AI and Data Intern Northeastern x IBM Boston, USA	Sep 2025 – Dec 2025
<ul style="list-style-type: none">Built RAG chatbot prototype using IBM watsonx.ai, Milvus, and Granite LLM for university capstone projectConfigured Milvus vector database with 768-dimensional embeddings and optimized Python retrieval pipelineCollaborated with Watsonx mentors to develop RAG chatbot for project management course materials	
Data Science Intern	April 2025 – July 2025
CommandL / Remote	
<ul style="list-style-type: none">Analyzed learner session logs to identify vague instructions and improve content clarity across training modulesAudited AI-generated content for tone, grammar, and logical consistency using structured evaluation rubricsBuilt LLM evaluation workflows using custom scoring metrics to guide prompt optimization	
AI Engineer	January 2025 – April 2025

CareEscapes AI Boston, USA	
<ul style="list-style-type: none">Designed and deployed healthcare chatbot using Azure OpenAI and LangGraph with FastAPI backend & PostgreSQLBuilt WebSocket and Redis infrastructure for real-time chat handling, state management and concurrent sessionsImplemented system health monitoring, retry logic, and automated session cleanup to improve system reliabilityCollaborated with cross-functional teams to build APIs and data models for consistent information flow	

Projects

Git-Guide - An AI Tutor for GitHub Repositories ([Live](#))

- Built a full-stack AI platform (**Next.js**, **FastAPI**, **PostgreSQL**) that converts repos into AI-guided learning journeys
- Implemented context-aware AI assistant using **LangChain** to generate adaptive coding tasks based on user skill level
- Optimized repo parsing & push verification with GitHub API, reducing processing latency by 40%
- Deployed on **Vercel** and **Render** serving active users with PostgreSQL storage and async processing

Song Popularity Prediction – End-to-End ML System ([Live](#))

- Engineered 13-step **Zen ML** pipeline with **ML flow** to process 600K+ tracks via efficient sampling strategies
- Extracted 18 audio features with Essentia and tuned 5 ML models (**XG Boost**, **Light GBM**), achieving $R^2 = 0.51$
- Deployed on Hugging Face with **Gradio** backend providing real-time predictions

SubsGen – AI Video Subtitle Generator ([Live](#))

- Implemented a production-grade video processing pipeline with Whisper AI, FFmpeg, FastAPI, and Next.js
- Designed a word-level subtitle timing and rendering pipeline aligned with Whisper token timestamps
- Deployed on Vercel and Hugging Face Spaces with a fully managed, serverless, zero-maintenance architecture

Technical Skills and Certifications

Core & Machine Learning: Python, SQL, PyTorch, TensorFlow, scikit-learn, XGBoost, LightGBM, Shell Scripting

AI & LLM Tools: Azure OpenAI, Hugging Face, LangChain, watsonx.ai, Milvus, Granite LLM, prompt engineering

ML Engineering: MLflow, ZenML, experiment tracking, feature engineering, ETL workflows, data validation

Backend & Infrastructure: FastAPI, REST APIs, WebSocket, PostgreSQL, Redis, Docker, Git/GitHub, PyTest

Databricks Certified ML Engineer Associate: Proficient in MLOps, Spark ML, Delta Lake, and distributed training

Microsoft Certified: Skilled in designing and deploying AI solutions using Azure OpenAI, Cognitive Services, Azure ML

Vaishnav Technologies: Machine Learning with Python Certificate