

# HARSHIT SONI

Bhanpura, Madhya Pradesh

+91 8458955735

[harshitsoni2026@gmail.com](mailto:harshitsoni2026@gmail.com)

[/harshitsoni29](https://www.linkedin.com/in/harshitsoni29)

[/Harshitsoni294](https://github.com/Harshitsoni294)

[Portfolio](#)

## Professional Summary

AI/ML Engineer and Computer Science student at IIIT Vadodara with deep expertise in neural networks, deep learning, and cloud-native machine learning. Engineered and deployed 8+ production-grade AI systems using PyTorch, TensorFlow, Scikit-learn, LangChain, and transformers, delivering up to 95% accuracy gains. Led hackathons, workshops, and competitive programming contests as a **technical committee member**, while also bringing strong stage presence as **Literary Club Secretary** and event host.

## Education

Indian Institute of Information Technology Vadodara - Diu Campus

2022 – Present

B.Tech (Computer Science and Engineering)

Diu, India

## Experience

Microsoft Azure AI Intern

May 2025

Microsoft + Edunet

- Completed 40+ hours of Azure AI training via MS Learn, completing 15+ modules covering GenAI and neural networks.
- Engineered content moderation system using Azure Cognitive Services with 95% accuracy across 10+ languages.
- Deployed serverless React application with Azure Functions, achieving sub-second response times.

Project Intern

June 2025 – July 2025

Acxiom Technologies

- Architected RAG system using LangChain and ChromaDB, processing 10,000+ documents with advanced chunking.
- Integrated LLMs with context-aware retrieval, boosting team productivity by 60% and cutting search time by 75%.
- Launched FastAPI service with analytics features, decreasing duplicate queries by 25% across teams.

## Technical Skills

- Programming Languages:** Python, JavaScript, TypeScript, C++, Java, SQL.
- AI & Machine Learning:** PyTorch, TensorFlow, LangChain, Hugging Face, ChromaDB, OpenCV, Scikit-learn.
- Web Technologies:** Next.js, React, Node.js, FastAPI, WebSockets, WebRTC, REST APIs.
- Tools & Databases:** Docker, Redis, MongoDB, MySQL, Git, AWS, Azure, Google Cloud.

## Projects

Speech Synthesis: Neural voice cloning System | Tacotron 2, HiFi-GAN, Gradio [Live Demo](#) December 2024

- Developed end-to-end TTS pipeline combining Tacotron 2 and HiFi-GAN models, training on custom voice dataset of 200 sentences to clone own voice.
- Achieved 4.2/5 MOS score through spectrogram enhancement and prosody modeling, delivering natural-sounding audio.
- Accelerated inference speed by 45% using optimized PyTorch pipelines and memory management, serving 500+ users via Gradio interface on Hugging Face Spaces.

ChessMasters: Multiplayer Chess Platform | Next.js, TypeScript, WebSockets, Redis [Live Demo](#) November 2024

- Created real-time multiplayer chess platform using WebSockets and TypeScript, supporting 200+ concurrent games.
- Implemented WebRTC video chat with Redis caching, achieving 99.9% uptime and sub-100ms latency.
- Optimized Kubernetes deployment with Docker and DigitalOcean, reducing infrastructure costs by 25% while maintaining high availability.

SentiX.AI: Smart Sentiment Analyzer | React, FastAPI, Gemini [Live Demo](#)

November 2023

- Extracted 500+ company-related posts and comments from Reddit API, designing React frontend with FastAPI backend architecture.
- Classified sentiment data with 92% accuracy, categorized content into positive, negative, and neutral.
- Automated PDF report generation for sentiment analysis trends, reducing manual processing time by 70% for business insights.

## Technical Achievements and Certifications

- Completed 8 hours NVIDIA Deep Learning Fundamentals course, mastering neural networks, CNNs, and model optimization. [Certificate](#)
- Achieved 4-Star rating on CodeChef with Current rating: 1885 and global rank: 3748. [Profile](#)
- Secured 3rd place among 200+ participants in Code Voyage 2024 (Inter-IIIT Coding Contest) by solving DP and graph theory problems. [Certificate](#)