Shreyas Makwana

Software Developer • Rajkot, Gujarat 360004

(+91)9510087633 shreyasmakwana.smh@gmail.com LinkedIn/in/Shreyas Github/Shreyas Portfolio

Education

Indian Institute of Information Technology Vadodara **August 2022 – June 2026**

B.Tech in Computer Science and Engineering

Work Experience

Tilva Artsoft **May 2025 – July 2025**

AI Software Engineer Intern *On Site — [Certificate]*

- Developed **FastAPI backend** and **Next.js frontend scaffolder** enabling one-click static deployments for client projects, reducing setup time by **80%**
- Automated **CI/CD** with **Docker**, cutting deployment time from **20 minutes to 4 minutes** and improving team velocity
- Implemented **Redis caching** and observability (**Prometheus, Grafana**), lowering **P95 latency to 1.8s** and increasing stability by **25%**

Projects

DocSense AI: FullStack RAG Intelligence Platform | *Next.js, FastAPI, PostgreSQL, Vector DB* [Live Link] **July 2025**

- Built cloud-deployed **RAG platform** with per-user **long-term memory** and **JWT-secured** accounts; ingests ≤**10MB PDFs** in ≈**6–15s**, preserving **25+** conversational turns across **300 validation runs**
- Engineered **RAG pipeline** using sentence-transformer embeddings and **pgvector**, achieving **92% citation accuracy**; average RAG response **3–5s** with vector search latency ∼**50–120ms** under realistic load
- Deployed cloud stack on Vercel/Render with Supabase Auth and **RLS** for data isolation; integrated **HF Router** (Llama-3.1-8B), **WebSocket**-ready APIs, **CI/CD** and tests, **2–3s** non-RAG latency

Sync Pad - Real-time Collaborative Text Editor | *Django, WebSockets, Yjs, React* [GitHub] **May 2025**

- Engineered **CRDT-based** collaborative editor (**Yjs, Remirror**) supporting **50+ concurrent users** with median update latency **<150ms** under simulated stress-testing
- Implemented **Django Channels** with **WebSockets** and **Redis**, improving synchronization efficiency by **65%** and cutting reconnection delays by **45%** under stress tests
- Containerized full-stack (**Django + Vite React**) with **CI/CD**; optimized caching and connection pooling to sustain **99.9% availability** and median API latency **<200ms**

Neural Text-to-Speech System | *Python, PyTorch, CUDA, Streamlit* [Live Link] **March 2025**

- Developed **Tacotron2+HiFi-GAN TTS pipeline** achieving **4.2/5 MOS** via spectrogram and prosody enhancement techniques for naturalness
- Accelerated **PyTorch** inference with **CUDA**-based data augmentation, reducing latency **45%** from **2.1s to 1.15s**
- Devised **Transformer grapheme to phoneme converter** with **98% accuracy** and scalable **Streamlit web UI**

Technical Skills

- **Programming Languages:** Python, C++, JavaScript, TypeScript, C, HTML, CSS, SQL
- **Machine Learning & AI:** PyTorch, TensorFlow, Scikit-learn, Keras, OpenCV, Transformers, CUDA
- **Web Technologies:** Next.js, React, Django, FastAPI, Node.js, Express.js, Streamlit, RESTful APIs
- **Databases & DevOps:** PostgreSQL, Redis, PGvector, Docker, Git, WebSockets, Prometheus, Grafana

Achievements

- Attained **3-Star** rating on CodeChef (**Current Ranking: 1687**) and solved **200+** problems on LeetCode with competitive **Rating of 1600** focusing on programming and DP [CodeChef] [LeetCode]
- Won **2nd place** in campus-wide Hackathon for developing a prediction-based **ML** solution with **91% accuracy**
- Earned **NVIDIA Deep Learning** Fundamentals certification mastering neural networks [Certificate]