

RISHIKESH

AI Research Engineer | Voice AI, LLMs & Agentic System

📍 Noida, India | 📞 +91 96501 84220 | rishikksh20@gmail.com | github.com/rishikksh20

PROFESSIONAL SUMMARY

AI Research Engineer with 10+ years in software development and 8+ years in AI and generative systems. Specializes in real-time Voice AI, multilingual speech, LLM architectures, RAG, multi-agent systems, and model post-training. Builds research-driven AI systems with a focus on efficient inference and production deployment.

PROFESSIONAL EXPERIENCE

MyClone.is (formerly Rappo)

Sept 2025 - 6-month engagement

AI Consultant / Developer | Remote (San Francisco-based)

- Developed a **real-time Voice AI digital persona platform** that reproduces users' voice, conversational style, knowledge, and communication patterns from text and audio.
- Built **low-latency LiveKit voice pipelines** integrating streaming STT/TTS, LLM orchestration, agent tools, and real-time session management.
- Implemented **persona-aware RAG and persistent memory** with LlamaIndex, PostgreSQL/pgvector, hybrid retrieval, document ingestion, and access filters.
- Migrated embeddings from **1,536 to 512 dimensions**, reducing vector storage **~3×**, improving retrieval speed **2-2.5×**, and cutting voice-chat latency **~15-20%**.

Open source: github.com/myclone-dev/myclone

Dubpro.ai (DeepSync Technologies)

Jan 2019 - Present

Co-founder & CTO | Noida, India

- Co-founded and led **AI dubbing, multilingual TTS, voice cloning**, speech transformation, and generative audio development.
- Architected **agentic RAG dubbing workflows** with human-in-the-loop review, increasing dubbing speed by **~2×**.
- Built an **in-house dubbing harness and LLM-as-judge quality verifier** to automate dubbing and QA, reduce manual review, and save **~40% in time** and **~25% in costs**.
- Developed **multilingual foundation-model capabilities** and applied **RLHF-based post-training** for translation quality and alignment.

Boutiqaat

Oct 2018 - Jan 2019

Data Science / Software Development Engineer | Gurugram, India

- Built **text-and-image product matching, hybrid recommendations**, and automated categorization for large e-commerce catalogs.
- Developed **clustering, tagging, and product-labeling algorithms** for commercial data operations.

Humonics Global

Dec 2017 - Oct 2018

Data Scientist | Gurugram, India

- Built **automated question-answer generation** and an **LSTM/WaveNet-based TTS engine** for conversational applications.
- Developed **face recognition and liveness detection** solutions.

Nucleus Software

Jun 2015 - Dec 2017

Software Engineer | India

- Developed **banking transaction analytics and loan-default prediction** systems.
- Built **financial supply-chain management software** using MVC architecture.

SELECTED OPEN-SOURCE PROJECTS

- **Nexus Code Agent** - CLI-first AI coding agent with MCP tools, skills, multi-agent orchestration, session/context state, persistent memory, and sandboxed agent workflows. github.com/rishikksh20/nexus-code-agent
- **iSTFTNet** - Lightweight mel-to-waveform synthesis combining learned upsampling with inverse short-time Fourier transforms. github.com/rishikksh20/iSTFTNet-pytorch
- **MyClone** - Open-source real-time digital persona and voice AI backend using LiveKit, LlamaIndex, FastAPI, PostgreSQL/pgvector, RAG, and agent orchestration. github.com/myclone-dev/myclone

TECHNICAL SKILLS

Models & Post-training: PyTorch, Hugging Face, Transformers, SSMS/Mamba, SFT, RLHF/RLVR, GRPO, distillation

Agents & Retrieval: MCP, RAG, multi-agent workflows, tool calling, LlamaIndex, LangChain, pgvector, persistent memory

Speech & Voice AI: TTS, STT, voice cloning, neural vocoders/codecs, LiveKit, Pipecat, Twilio

GPU & Inference: Triton, CUDA, vLLM, FlashAttention, KV cache, quantization, GPU memory and latency optimization

Backend & Cloud: Python, FastAPI, async programming, SQL, PostgreSQL, Redis, Docker, AWS, GCP

EDUCATION

Institute: National Institute of Technology, Silchar, India

Degree: Bachelor of Technology (B.Tech)

Branch: Electronics & Instrumentation Engineering

Years: 2011–2015

LANGUAGES

- Hindi (native)
- English (proficient)