

Chirag Gupta

chirag.gupta.290403@gmail.com | +91 7827721664

linkedin.com/in/chiruu12 | github.com/chiruu12 | chiruu12.github.io

AI engineer building multi-agent systems and agent infrastructure. GSoC '25 contributor and '26 mentor at Jenkins.

PROFESSIONAL EXPERIENCE

AI Engineer, GoManzanas (US Startup)

May 2025 - Present

Remote

AEGIS, Multi-Agent Pipeline: Detected pirated CAD software usage by analyzing telemetry and signal patterns.

- Built multi-agent pipeline for pirated CAD detection: 89% precision (94% manual baseline), 95% cost reduction (\$1.06 to \$0.05/case)
- Built multi-backend LLM orchestration with fallback logic and stage-specific prompting

Ace1t, Real-Time Voice AI: Built voice-AI backend for a real-time accent training platform.

- Integrated multiple TTS models with ASR (OpenAI Whisper, FastConformer) for real-time voice processing
- Designed push-to-talk pipeline with streaming audio, model-driven speech control, and silence handling

Booster, AI Collaboration Platform: Built multi-agent infrastructure for an ideation platform (FastAPI + React + PostgreSQL).

- Built multi-agent group chat with turn-based orchestration and real-time collaboration across agent personas
- Designed containerized code execution environment and agent session lifecycle with idle cleanup and queue-based concurrency
- Built custom MCP server, skills system with model tiering, and agent benchmarking framework

GSoC '25 Contributor | '26 Mentor, Jenkins

May 2025 - Present

- Built modular multi-agent diagnostic system (log-parser, classifier, recommender) for Jenkins build failures
- Added vector-store RAG layer, benchmarked on 20 curated questions: 95% context relevance, 3.75/5.0 overall score
- Integrated multi-backend LLM adapter (OpenAI, Gemini, Claude) with mode-switching for analysis vs recommendation

AI Engineer, Reasonify Technology Pvt Ltd

- Built multi-agent LLM pipelines for adaptive learning content and automated study-plan generation
- Built automated Manim video generation pipeline (CLI + prompt standardization + component library), cutting production time from 5h to 20min
- Implemented hybrid semantic filtering for TTS using regex rules and a fine-tuned Gemma 270M model

Open Source

- Rago** (OSL Incubator): Added 5 API backends (Cohere, Fireworks, Together, Groq, HuggingFace) for embeddings and generation. 7 merged PRs.
- Keras:** Integrated `np.diff` and `np.log2` via OpenVINO backend.
- Agno:** Added Exa Answer method, migrated deprecated models across 7+ providers, URL expansion for Crawl4ai.

AI & Data Science Intern

May 2024 - Jan 2025

Mindcraft Labs | EvoAstra Ventures

- Mindcraft Labs: Automated 19+ workflows with RETELL, Voiceflow, and Make
- EvoAstra Ventures: Built fraud-detection model on 6.3M+ records, 96.5% accuracy, 94% fraud recall

PROJECTS

Hive: Local-first agent OS. Persistent agents from YAML, multi-model routing (Claude, Codex, LM Studio), agent rooms for collaboration. github.com/chiruu12/Hive

NexNet: Neural network framework from scratch in NumPy. BERT and GPT with multi-head attention and positional encoding. No PyTorch, no TensorFlow. github.com/chiruu12/NexNet

EDUCATION

M.Sc. (Hons.) Mathematics & B.E. (Hons.) Electronics & Instrumentation

2021 - 2026

BITS Pilani, Goa Campus, Goa, India

SKILLS

Languages: Python, C/C++, SQL

ML/AI: PyTorch, LangChain, vLLM, Ray, Hugging Face Transformers, MLFlow

Infrastructure: Docker, Git, FastAPI, Pydantic, MCP, gRPC

ACHIEVEMENTS

- Top 20 in Meta Hacker Cup Open AI practice round (13,000+ participants)
- Winner, Smart India Hackathon (SIH) 2025