

# GURNOOR SINGH SAINI

AI Engineer

📞 +91 94653 01124 ✉️ [sainigurnoor511@gmail.com](mailto:sainigurnoor511@gmail.com) 🌐 [github.com/Sainigurnoor511](https://github.com/Sainigurnoor511) 🌐 [linkedin.com/in/sainigurnoor511](https://www.linkedin.com/in/sainigurnoor511)

## Summary

AI Engineer with 2 years of production experience building real-time LLM systems, RAG pipelines, and agentic AI workflows at Jungleworks. Specializes in deploying scalable, domain-specific GenAI solutions using FastAPI, LangChain, and cloud infrastructure, driving measurable cost reduction and automation across enterprise platforms. Seeking AI/ML engineering roles focused on LLM applications and intelligent system design.

## Technical Skills

**Programming:** Python, Bash, MySQL, PostgreSQL, MongoDB

**Frameworks & Libraries:** FastAPI, Rest APIs, Flask, Next.js, LangChain, Streamlit, Autogen

**AI/ML:** Large Language Models (LLMs), Audio-to-Audio LLMs (Ultravox), Prompt Engineering, Agentic AI, Embeddings, Model Context Protocol (MCP), Retrieval-Augmented Generation (RAG), OpenAI API

**GenAI Techniques:** Chatbots, Time Series Forecasting, Vector Databases, Model Evaluation

**Tools & Platforms:** Git, Docker, Linux, Postman, Cursor, Claude Code, Github Copilot

## Work Experience

### Jungleworks

AI Engineer

June 2024 – Present

Chandigarh, India

- **Spearheaded Full-Stack Development of AI Dialer (Super SDR)** — a real-time conversational AI calling system for various business use cases using Fixie.ai's Ultravox model achieving **2,000x cost efficiency** over OpenAI's real-time model.
- **Integrated Generative AI Across Five Core Platforms** — automated workflows and personalized experiences in Bulbul (CRM), Tookan (Delivery Management), Hippo (Customer Engagement), Yelo (Hyperlocal Marketplace), and Outplay (Sales Automation), reducing manual effort by **40%** and increasing user engagement by **25%**.
- **Engineered AI Chatbot Generator for Hippo** — convert helpdesk data into deployable LLM chatbots within 2 hours; deployed across Tookan, Outplay, and client sites, reducing support load by **70%**.
- **Developed Forecast Engine for Jugnoo** — time-series model predicted daily rides and rider availability with **98% accuracy**, optimizing planning and resource allocation.
- **Outplay AI Initiatives** — built Ideal Customer Profile detection, lead scoring, and reply generation pipelines, increasing qualified lead flow by **30–40%**.

## Projects

**Super SDR** | [supersdr.ai/ai-dialer](https://supersdr.ai/ai-dialer) | FastAPI (Python), Next.js, Llama 3.2 7b LLM, Ultravox, AWS S3

- Built AI Dialer and AI Sales Trainer, a real-time AI sales caller and trainer (FastAPI + Next.js) handling **50+ concurrent calls** with sub-second latency.
- Fine-tuned Ultravox for domain-specific calls, achieving **2,000x cost savings** over OpenAI's real-time model.
- Enabled multi-voice agents with customized gender control, cutting deployment time by **80%**.
- Integrated lead management, transcripts, and call recordings, improving follow-up efficiency by **35%**.

**RAG Chatbot** | [github.com/Sainigurnoor511/rag-chatbot](https://github.com/Sainigurnoor511/rag-chatbot) | Python (FastAPI), LangChain, ONNX, Redis, Chroma, Hugging Face

- Converted model to ONNX format for local inference, reducing embedding time by **50%** and overall latency by **40%**.
- Implemented Chroma as a vector store, enabling fast retrieval from **100K+ document embeddings**.
- Applied Redis caching, improving session management performance by **50%**.

**Artify.ai** | [artify-ai.streamlit.app](https://artify-ai.streamlit.app) | Python, Stable Diffusion (Flux Schnell), Streamlit, CI/CD

- Built a Streamlit-based image generation app using the Flux Schnell API, deployed on GitHub with CI/CD for automated updates.
- Designed an intuitive generation interface enabling custom prompt-based image creation.

## Education

### GNA University, Phagwara

B.Tech in Computer Science and Engineering — CGPA: 8.5+

2021 – 2025

Punjab, India

## Certifications

---

- IBM Certified Full-Stack Developer
- AWS Academy Data Center Technician
- Application Development using Microservices & Serverless (IBM)
- Node.js & MongoDB: Back-end Development (IBM)

## Achievements

---

- Won **Best Innovation Idea & Pitching Award** twice in competitive national hackathons.
- **Top 10** finalist at HACKMOL 5.0, National Institute of Technology Jalandhar.
- Google Cloud Skill Boost Badges: Cloud Fundamentals, Core Infrastructure, Core Services.