

# Sunil Giri

📍 Delhi, India    📁 Portfolio    🔗 LinkedIn    🐙 Github    ✉ seungiri841@gmail.com    ☎ 8218452987    💬 Skype

AI/ML Engineer with 1 year of experience in designing, developing, and deploying scalable machine learning models and AI-driven applications. Proficient in Python, Django, TensorFlow, NLTK, and Scikit-Learn with a strong track record of delivering measurable business outcomes. Skilled in integrating AI/ML solutions with business requirements, leading cross-functional teams, and driving innovation in real-world projects. Recognized for contributions to impactful AI products.

## WORK EXPERIENCE

- AI/ML Developer**, Mobiloitte Technologies India Pvt. Ltd. 📄 June 2024 – March 2025 | Delhi, India
- Designed and deployed scalable **AI/ML** architectures, achieving a 95% **accuracy** in **predictive** analytics models.
  - Spearheaded 5+ **pre-sales** engagements, **contributing** to a 15% growth in **client acquisition** and **retention**
  - **Spearheaded** the end-to-end development of **products** for 5+ **clients**, implementing **scalable** backend APIs with efficient **load balancing** and **automation** solutions.
  - **Optimized** AI models, **improved** model **performance** by 30%
  - **Developed** and managed end-to-end **pipelines**, **containerized** entire projects using **Docker**, **Compose**, and **Jenkins**, achieving a 90% automation rate in project **deployment**.
- AI/ML Engineer**, Presage Insights 📄 March 2025 – present | Noida, India
- **Designed and deployed a scalable RAG-based asset-specific chatbot** using Django, enabling document uploads, conversation context tracking, and integrated web search to deliver accurate, curated insights for industrial equipment diagnostics.
  - **Built RESTful APIs for LLM interaction and document ingestion**, leveraging PostgreSQL for data storage and Pinecone for high-performance vector embedding of asset-specific documents; with real-time data insights summarization.
  - **Led end-to-end containerized deployment on AWS** using Docker, Nginx (reverse proxy), Celery with Redis for asynchronous task management, and DevOps tools like PuTTY and FileZilla, ensuring high availability and robust performance in production.

## EDUCATION

- Bachelor of Computer Application**, Tula's Institute 📄 2021 – 2024 | Dehradun, India  
Percentage: 78%
- Master of Computer Applications**, Inmantec Institutions 📄 2024 – 2026 | Ghaziabad, India  
Relevant Coursework: ML & AI, Data Structures & Algorithms, DBMS, Cloud Computing & Distributed Systems

## KEY SKILLS

Languages:	Cloud & Database	Data Science
Python(Pandas, NumPy.), C/C++, Excel	SQL, Postgres, Celery, Docker, Compose,	Machine Learning, NLP, Gen AI,
<b>Frameworks:</b> Django, DRF, Flask,	Nginx, Jenkins, Redis, Prometheus,	Langchain,LLMs, RAG, Agno, Scikit-learn,
TensorFlow, Web Scraping, Matplotlib	GitLab, CI/CD, Postman, AWS EC2, PuTTY,	NLTK, Transformers, Matplotlib, Hugging
<b>APIs:</b> RESTful APIs, DRF, WebSocket	FileZilla, S3, PowerBI	Face, Statistics

## PROJECTS

- ICE-BUTTON(SaaS)**, (Twilio, Raspberry Pi, Django, RTMP Protocol)
- **Secure IoT Video Streaming:** Developed a secure server on **Raspberry Pi** using **Nginx** to enable **RTMP protocol**-based video streaming from IoT devices, **integrated** with HTTP for **seamless** browser-based viewing.
  - **Scalable Backend Architecture:** Improved and implemented robust **Django-based APIs** for admin **functionalities**, **payment gateway integration**, communication features, and **subdomain management**—ensuring efficient **multi-tenancy** for the **SaaS** platform.
- AI Video Generator**, Web Scraping, Automation, Text-to-speech
- **Dynamic Content Extraction:** Built a Python-based **backend** using web **scraping** frameworks (BeautifulSoup, **Selenium**) to **extract** content from 100+ web pages, **automate** scrolling, and merge website data with audio assets and **text-to-speech** intros.
  - **Optimized Video Processing:** Operated load balancing and **asynchronous** processing using tools such as **Celery/AsyncIO**, reducing video **generation** time by up to 40% for high-volume data inputs.
  - **Automated Template Integration:** Proposed robust **ETL pipelines** to process and **validate** user-defined templates, **automating** **personalized** video outputs with **precise** content and audio **alignment**.
- Chicken Disease Classifier Using MLOPS DVC Pipeline, ML & AWS** 📄
- **CNN-based Disease Prediction:** Crafted a CNN model using **TensorFlow** and Python that achieved 93% **accuracy** across 5+ common chicken **diseases** while reducing **detection** time by 80%.
  - **MLOps Pipeline:** Applied an end-to-end MLOps pipeline with **DVC** and **GitHub Actions** to **automate** model versioning and **deployment** processes—resulting in 65% **faster** iteration cycles and 99% **reproducibility**.
- QNA BOT : Using Langchain-RAG**, GenAI, HuggingFace, VectorDB, Faiss, LLM 📄 January 2025 – February 2025
- **AI Assistant Development:** Devised a **full-stack AI Assistant** using Python, **Socket.IO** 📄, and **JavaScript** that **processes** 15+ document formats with 90% **accuracy**, reducing **analysis** time by 75% and enabling **real-time** responses within 2–3 seconds.
  - **Memory Retention System:** Architected a system supporting 100+ **concurrent** users with 98.9% uptime by **implementing** **WebSocket** connections to handle 1000+ messages/minute while **maintaining** **contextual** accuracy above 95%.
  - **Responsive Web Interface:** Defined a **responsive interface** for 50+ **enterprise** users that **integrates** **RESTful APIs** and **Socket.IO** 📄 **events** to process 200+ file uploads/hour, **achieving** 40% faster response times through **optimized** frontend-backend communication.

## CERTIFICATES

Machine Learning with Python - FreeCodeCamp 📄 | AWS Academy Graduate | AI/ML Developer 📄