

ARUN SANYAL

✉ arunsanyals@gmail.com

☎ +1(520)256-6448

📍 Tucson, AZ

in [linkedin.com/in/arun-sanyal/](https://www.linkedin.com/in/arun-sanyal/)

🐙 github.com/ArunSanyal

Education

University of Arizona, Tucson

Master of Science in Computer Science

Courses: Design Analysis of Algorithms, Principles of Computer Networking, Software Engineering

Tucson, Arizona

(2025 - 2027)

Fr. Conceicao Rodrigues College of Engineering

Bachelor of Engineering (Honours in Blockchain) - Computer Engineering : CGPA: 8.66

Courses: Data Structures and Algorithms, Databases, Computer Networks, Operating System, etc

Bandra, Mumbai

(2021 - 2025)

Skills Summary

Languages: Python, Java, C, JavaScript, TypeScript, SQL, Dart, Bash

Technologies / Frameworks: React, Tailwind CSS, Bootstrap, HTML/CSS, Flutter, Vite, Firebase, Android, Unreal Engine 5

Machine Learning: PyTorch, TensorFlow, HuggingFace Transformers, Scikit-learn, OpenCV, OCR (Tesseract/PaddleOCR), LangChain, LangGraph, RAG Pipelines, Streamlit, Pandas, NumPy, Matplotlib

DevOps: AWS (EC2, S3, Lambda), Docker, Git, GitHub, Vercel, Firebase Hosting, HPC (Slurm, multi-GPU training for Qwen, LLaMA, and Sentence Transformers)

Tools & Platforms: VSCode, JupyterLab, Postman, PgAdmin, Linux, Windows, Arduino, Kaggle

Work Experience

GenAI R&D Intern – Software/ML Intern – Konsultera Solutions Pvt. Ltd.

Feb 2025 - July 2025 (Onsite)

- Built LLM pipelines for field extraction from 500+ PDFs using LangChain, multi-OCR, and schema-driven prompts, cutting manual effort by ~60% and improving cross-team communication.
- Deployed LangGraph agents + knowledge graphs to automate workflows (validation, dataset creation, email automation), achieving 50–60% accuracy gains via optimized prompts and Pandas pipelines.
- Fine-tuned Gemini, Qwen-2B, LLaMA-3.2/8B, Pic2Struct on DocVQA-scale datasets (~1,200 folders × 7 PDFs), improving accuracy, reducing latency, and guiding final model selection.
- Leveraged HPC resources (NVIDIA RTX A1000 + RTX 4090 GPUs) with multi-threading to achieve 3× faster training and inference throughput; delivered Streamlit PoCs that transformed PDFs into structured JSON/CSV outputs + cropped images, later scaled into enterprise adoption; integrated HuggingFace Transformers and LiteLLM routing for large-scale, production-ready deployment.

Projects

HerbalVista – Immersive 3D Educational Game on Ayurvedic Plants (Group of 3)

July 2024 - May 2025

- Unreal Engine 5, RealityCapture, ConvAI, RAG (LLM), React, TypeScript, Vite, 3D Modeling, Photogrammetry
- Developed a photorealistic 3D herbal garden to teach children Ayurvedic plants, using photogrammetry-based high-fidelity reconstructions and Unreal Engine 5 for immersive gameplay.
- Integrated ConvAI agents + RAG chatbot (HuggingFace, LangChain) for multilingual Q&A and contextual learning; companion React + Vite web app with bookmarks and real-time knowledge assessments.

AI-Powered Food Safety Detection System

Jan 2024 - May 2024

- TensorFlow, MobileNetV2, OpenCV, OCR, Gemini API, Streamlit, Plotly, Python, JupyterLab
- Engineered an AI-based system that scans food packaging/manual inputs to detect harmful ingredients & allergens using MobileNetV2 + OCR, achieving high accuracy on test datasets.
- Integrated Gemini API for predictive health-risk assessment (e.g., diabetes, heart disease) and Streamlit dashboards with Plotly for real-time visualizations, boosting food safety awareness.

LLM-driven Product Extraction & Cataloguing

Jan 2025 - Mar2025

- Python, Streamlit, LangChain, HuggingFace, pdf2image, Pillow, Tavily Search
- Built a 2-stage GenAI pipeline (Extractor + Cataloguer) converting supplier brochure/catalog PDFs into structured SKU spreadsheets and per-SKU image crops, with strict JSON schema + precision filters yielding 90%+ clean outputs.
- Automated web enrichment via Tavily (descriptions, images, source URLs) with domain filters, and shipped an interactive Streamlit UI—preview/download, per-PDF workspaces, reproducible folders—container-ready, scalable for enterprise use..

COURSES AND CERTIFICATIONS

- Building Agentic RAG with LlamaIndex | [DeepLearning.AI](#)
- AWS Academy Cloud Foundations | [AWS Academy](#)
- Ai Agents in LangGraph | [DeepLearning.AI](#)
- The Complete Flutter Development Bootcamp with Dart | [Udemy](#)