

# Aniket Surjuse

Pune, MH | +919511875643 | asurjuse111@gmail.com | [aniketsurjuse.github.io](https://aniketsurjuse.github.io) | [github.com/AniketSurjuse](https://github.com/AniketSurjuse)

## EXPERIENCE

---

### Siemens Energy – Pune, MH

AI/ML Developer

July 2024 – Present

- Developed a GPT-4-powered internal QA chatbot, reducing engineering teams' information retrieval time by **80%**.
- Built backend APIs using **FastAPI** and integrated **CosmosDB** for vector search to support RAG-based LLM responses.
- Automated AWS infrastructure (S3, Lambda, RDS) using **CDK**, improving deployment speed and reducing manual provisioning.
- Contributed to AI solution to auto-generate comments on customer documents by fine-tuning **T5**, **LLaMA 2**, and **DistilBERT**, cutting review time by **90%**.
- Applied **LoRA**, **QLoRA**, and **PEFT** techniques using **Azure ML** to optimize model training for cost-efficiency.
- Integrated model evaluation steps to ensure response quality and compliance with internal standards.

AI/ML Intern

Jan 2024– June 2024

- Built an automated certificate validation system using **Google Document AI OCR** to extract key chemical and mechanical properties.
- Designed an LLM-driven validation agent (via **LangChain**) that performed unit conversions, flagged compliance gaps, and achieved over **95% accuracy**.
- Developed an automated reporting pipeline that generates compliance reports with deviations highlighted, reducing review time from **hours to minutes**.

## SKILLS

---

**Languages:** Python, C++, SQL

**AI/ML:** Pytorch, LangChain, LangGraph, Scikit-learn

**Frameworks:** FastAPI, Streamlit

**Cloud & Tools:** AWS (S3, Lambda, Sagemaker, CDK), Azure ML, Git, Docker

**Core Expertise:** LLMs, RAG, Agentic AI, NLP

## PROJECTS

---

### AGENTIC DOCUMENT EXTRACTION ([source code: https://github.com/AniketSurjuse/agent-pdf-extractor](https://github.com/AniketSurjuse/agent-pdf-extractor))

- Developed an AI-powered PDF extraction tool using **LangGraph** (for agent workflows) and **Pydantic** (for schema validation) with **Azure OpenAI** LLM.
- Extracts and preserves **document structure** - text, tables, and images and outputs clean, well-formatted **Markdown**.
- Built modular **document parsers** with fallback strategies using LLM-driven reasoning and tool selection.

### GPT2 FROM SCRATCH ([Live Demo](#))

- Rebuilt GPT-2 architecture from scratch in PyTorch to understand self-attention mechanisms, tokenization, and causal language modeling.
- Initialized from Hugging Face's 355M-parameter GPT-2 checkpoint and fine-tuned on instruction-following data using causal language modeling.
- Converted model to float16, reducing size by 50% for faster CPU inference and deployment.

## EDUCATION

---

### GOVERNMENT COLLEGE OF ENGINEERING, AURANGABAD

June 2024

Bachelor of Technology, Computer Science; Cumulative GPA: 9.02/10

## CERTIFICATIONS

---

- **Complete Agentic AI Bootcamp With LangGraph and Langchain** - Udemy
- **Deep Learning Specialization** - Coursera (Andrew Ng)
- **Machine Learning Specialization** - Coursera (Andrew Ng)