SAYAK DUTTA

Github: github.com/sayakdutta | E-Mail ID: **sduttan598@gmail.com** | LinkedIn: **Sayak.Dutta**

Mobile Number: **(+49) 17679861928** | Address: **Munich (Open to Relocation - London or EU)**

Senior Machine Learning Engineer with 9+ years of experience designing and scaling production-grade AI systems. Specialized in **LLMs, RAG pipelines, real-time analytics, and full-stack AI platforms**, delivering end-to-end solutions that are **scalable, low-latency, and production-ready**. Skilled in **distributed infrastructure, MLOps pipelines, and intelligent applications** across Python, C++, TypeScript (Node.js, Next.js), Docker/Kubernetes, and modern cloud stacks. Thrive in **high-ownership, fast-paced environments** where rigorous engineering meets applied machine learning.

CORE COMPETENCIES

Programming: Python (Advanced), C++, TypeScript/JavaScript (Node.js, Next.js), LangChain, OpenAI, Hugging Face
ML/AI Techniques: Time-series forecasting, NLP, recommendation engines, LLMs, deep learning, feature engineering, model validation
Data Engineering: ETL pipelines, time-series & structured data modeling, SQL (PostgreSQL)
Distributed Systems: Kafka, Redis, WebSockets, FAISS, Pinecone
Cloud & DevOps: Docker, Kubernetes, AWS (Lambda, S3), CI/CD (GitHub Actions)
ML Ops & Infra: Real-time ETL, model deployment, vector databases, pipeline orchestration
Other: FastAPI, Streamlit, Git/Github, Agile/Scrum, Open-Source contributions, Performance optimization

PROFESSIONAL EXPERIENCE

Senior Machine Learning Engineer, **Torto Labs Pvt Ltd.**, India ***Aug 2021 – Present***

- Architected and deployed a distributed **RAG knowledge assistant (LangChain, FAISS, Redis)** supporting 100+ daily users, reducing lookup latency by 60%.
- Built core **ETL infrastructure (Python, Docker, AWS Lambda)** to process **1M+ records/month**, powering ML-driven forecasting.
- Streamlined **CI/CD pipelines** via **Docker** and **GitHub Actions**, achieving **95% reduction** in manual deployment time.
- Designed real-time **dashboards (Streamlit, PostgreSQL)** for analytics, accelerating decision-making by **35%**.

Machine Learning Engineer, **Berylls GmbH**, Germany ***Jul 2023 – Dec 2023***

- Developed and productionized real-time **recommendation engines** (**XGBoost**, collaborative filtering), increasing engagement by **25%**.
- Created scalable **NLP/ML pipelines** to classify **100K+ customer queries/month**, delivering **90% tagging accuracy**.
- Automated **ML deployment workflows (Docker, GitHub Actions)**, reducing operational overhead by **80%**.
- Collaborated with research/product teams for rapid delivery of new **AI features** in agile sprints.

Senior Data Analyst, **Nailbiter Inc.**, India ***Oct 2020 - Feb 2021***

- Engineered **text analytics pipelines** using **Hugging Face Transformers** for **200+ hours** of data, enabling real-time semantic search.
- Automated **API data ingestion (FastAPI, PostgreSQL)**, optimizing campaign targeting by **20%**.
- Shortened ideation-to-deployment cycle by **30%** via agile collaboration with cross-functional teams.

Analyst, **eClerx Services Ltd.,** India ***Apr 2018 – Oct 2020***

- Designed and integrated a **fuzzy-matching microservice (Python, Redis)** processing **500K+ records**, boosting data accuracy by **35%**.
- Implemented **real-time monitoring (Kafka)** to maintain **99.9% uptime** for distributed pipelines.
- **Built compliance dashboards (Power BI)** reducing manual data retrieval by 40%.

Project Manager, **HTL Aircon Pvt Ltd.,** India ***Dec 2016 – Mar 2018***

- **Led cross-functional teams** to deliver 10+ technical projects with 95% on-time completion.
- **Digitized reporting workflows** with Python, reducing processing time by 4 hours/week.

PROJECTS

ResearchRAG - AI Research Paper Digestor & Explorer

- Engineered a **full-stack RAG system** (FastAPI + FAISS) that processes 30-page academic PDFs into summaries, strengths/weaknesses, and future research directions in under 30 seconds.
- Built an **LLM-powered interactive chat** grounded in embeddings, delivering 95%+ context-relevant answers using free OpenRouter GPT-OSS models.
- Delivered a **production-ready solution** with robust error handling, logging, and seamless frontend–backend integration.
- Designed a **responsive UI** with exportoptions (PDF/Markdown), making research digestion 3–5× faster for researchers and students.

SAYAK DUTTA

Github: github.com/sayakdutta | E-Mail ID: **sduttan598@gmail.com** | LinkedIn: **Sayak.Dutta**

Mobile Number: **(+49) 17679861928** | Address: **Munich (Open to Relocation - London or EU)**

Tech stack: FastAPI, Next.js, Tailwind, TypeScript, FAISS, PyMuPDF, OpenRouter GPT-OSS.

AskPostgres - Natural Language to SQL for PostgreSQL

- Built a full-stack application that converts natural language into SQL queries, enabling **business users to query PostgreSQL databases without technical SQL knowledge**.
- Implemented authentication, role-based access control, error handling, and interactive UI with light/dark themes; validated with **extensive query variations** to ensure robustness.
- Improved analyst efficiency by **reducing SQL-writing time by up to 80%**, supporting business analytics, data exploration, and performance monitoring.

Tech stack: FastAPI, Next.js, Tailwind, TypeScript, FAISS, PyMuPDF, OpenRouter GPT-OSS.

OutreachPilot - Campaign Automation & Tracking Platform

- Built a **full-stack automation platform** (FastAPI, Next.js, Tailwind, PostgreSQL, Docker) for seamless campaign orchestration and monitoring.
- Integrated **automated email delivery** via Resend API with custom **pixel + redirect tracking** to measure open rates, CTR, and variant performance.
- Designed modular **backend microservices** with async workers, queues, and real-time logging to ensure scalability and fault tolerance.
- Developed a **responsive dashboard** with secure authentication, role-based access, and light/dark modes, enabling non-technical users to launch and track campaigns easily.
- Delivered **production-ready infra** with containerized deployment (Docker) and CI/CD pipelines, showcasing enterprise-grade reliability and scalability.

Tech stack: FastAPI, Next.js, Tailwind, TypeScript, PostgreSQL, Docker, Resend API

TalkToTube - YouTube Transcript Explorer with LLM Q&Arer

- Built a **full-stack application** (FastAPI, Next.js, Tailwind, PostgreSQL, Docker) that ingests and processes YouTube transcripts into searchable, structured text.
- Implemented an **LLM-powered Q&A system** enabling users to ask natural language questions and receive context-grounded answers from video content.
- Designed a **responsive web interface** with role-based access, real-time results, and export options for summaries and highlights.
- Integrated **vector search (FAISS)** and embeddings for efficient retrieval, delivering fast and accurate results across long-form video content.
- Delivered production-ready infra with **containerized deployment (Docker)** and CI/CD pipelines, ensuring scalability and reproducibility.

Tech stack: FastAPI, Next.js, Tailwind, TypeScript, PostgreSQL, FAISS, Docker

EDUCATION

University of Mumbai, Mumbai, India — ***August 2012 – May 2016***

Bachelor of Engineering– Mechanical Engineering | Grade: First Division

Languages: English (Fluent), Hindi (Native), German (A2)

Visa: Valid in Germany; open to relocation