

Sagar Sikdar

Backend / Data Engineer | ML LLM Pipelines

Nürnberg, Germany | +49 176 7610 2783 | sagarsikdar2812@gmail.com

GitHub: github.com/sagar-crypto | LinkedIn: linkedin.com/in/sagarsikdar

Experience

Working Student — Siemens Healthineers

Jan 2024

Erlangen, Germany

Present

- Built automated workflows and integrated systems via SharePoint/Power Platform; delivered in an enterprise environment.
- Built ETL-style extraction + preprocessing pipelines and visualized KPIs in Power BI for stakeholders.
- Integrated workflows into SharePoint-based systems; improved maintainability with reusable scripts/components.
- Tech: Power Automate, Power Apps, Power BI, SharePoint, PowerShell, C#/ .NET

Software Engineer — Edmingle

Jul 2022

Bangalore, India

Sep 2023

- Designed and maintained Jenkins-based CI/CD pipelines and Dockerized services; reduced deployment time by 30%+.
- Automated cloud infrastructure and deployments on AWS (EC2, RDS, ElastiCache), improving reliability and scalability.
- Built REST APIs/tools and executed data migrations (8+ platforms); remediated **80%** of VAPT findings.
- Tech: PHP (Laravel), Node.js, SQL, AWS, Docker, Jenkins, Git

Software Engineering Intern — Edmingle

Jan 2022

Bangalore, India

Jul 2022

- Implemented a Node.js log-processing service and containerized internal dev tools.
- Improved automated testing and contributed deployment scripts and compliance tasks (GDPR/ISO).

Projects

Time-Series Anomaly Detection (M.Sc.)

Apr 2025

Germany

Oct 2025

- Designed end-to-end anomaly detection pipelines for high-frequency industrial and power-grid time series.
- Benchmarked Transformer-based and deep learning models using domain-relevant metrics (F1, FP/FN trade-offs).
- Implemented reproducible experimentation with MLflow and Optuna; enforced identical preprocessing for training and inference.
- Tech: Python, PyTorch, Hugging Face TimeSeriesTransformer, pandas, scikit-learn, MLflow, Optuna

Retrieval-Augmented Generation (RAG) System — In Progress

2025

Germany

Present

- Designing a Retrieval-Augmented Generation pipeline to answer domain-specific queries using external knowledge sources.
- Implementing document ingestion, chunking, and embedding workflows for structured and unstructured data.
- Building a semantic retrieval layer using vector similarity search to ground LLM responses in retrieved context.
- Tech: Python, LangChain/LlamaIndex, Vector Databases (FAISS/Chroma),

M.Sc. Thesis (In Progress)

Known Operator Learning for Fault Localization in Power Grids

Pattern Recognition Lab, FAU
Erlangen–Nürnberg

- Hybrid physics–ML approach for fault localization using distance protection models.
- Learning systematic correction terms on top of classical impedance-based estimators.
- Experiments on large-scale simulated time-series data from inverter-dominated grids.

Core Competencies

Backend & Data Pipelines:

Design and maintenance of robust backend data-processing pipelines for ML and LLM-based systems; ETL-style workflows, data validation, and preprocessing.

APIs & System Integration:

Backend services and integrations using REST APIs; structured request/response handling, database-backed services, and service-oriented architectures.

Machine Learning & LLM Workflows:

Preparation of datasets, experimentation, and evaluation of ML models; integration of LLMs into production workflows (e.g., RAG pipelines).

Cloud & DevOps:

CI/CD pipelines (Jenkins, Git-based workflows), containerization with Docker, cloud environments on AWS and Azure; working knowledge of Kubernetes.

Architecture & Code Quality:

Clean code principles, modular backend design, scalable systems, and maintainable architectures in agile development environments.

Collaboration &

Documentation: Structured collaboration with stakeholders and teams; clear documentation of

Embeddings, LLMs, FastAPI

Screwdriving Time-Series Clustering (M.Sc.)

Oct 2024

Germany

Mar 2025

- Analyzed multivariate industrial screwdriving curves (torque/angle/time) to identify process variability and hidden failure patterns.
- Applied VAE-based representation learning to compress high-dimensional time series into structured latent spaces.
- Compared multiple VAE architectures to study the impact of latent dimensionality and inductive biases.
- Performed exploratory data analysis to understand distribution shifts and noise characteristics across production runs.
- Clustered latent representations to detect process drifts, anomalies, and regime changes without explicit labels.
- Implemented reproducible training, logging, and hyperparameter optimization pipelines using MLflow and Optuna.
- Tech: Python, PyTorch, Variational Autoencoders, Optuna, MLflow

Hackathons

Wellness AI Multi-Agent Helper (2025)

- Designed an agentic LLM system to orchestrate wellness-related workflows across multiple sub-agents.
- Implemented real-time voice-to-voice interaction using a Python FastAPI backend and Kotlin Multiplatform mobile client.
- Implemented backend services using FastAPI and Pydantic for structured request/response validation and ML workflow integration.
- Tech: Python, FastAPI, Gemini, WebSockets, Kotlin Multiplatform, Android

Sustainability Reporting (2024)

- Built a supplier portal for ESG data entry and certificate uploads aligned with EUDR requirements.
- Implemented an LLM-based FAQ chatbot to reduce manual clarification and regulatory confusion.
- Tech: Django, React, GPT-3.5 Turbo, MySQL

Publications

A Study on Kerberos and Graphical Password-Based Web Authentication Scheme

LNNS, Springer, 2023

- Proposed and evaluated two authentication models combining Kerberos with graphical passwords.
- Analyzed security–performance trade-offs with and without Diffie–Hellman key exchange.

Education

M.Sc. Data Science — FAU Erlangen–Nürnberg

Oct 2023

Erlangen, Germany

Present

- Focus: AI, Pattern Recognition, Deep Learning, Machine Learning.

B.Tech. CSE (Cyber Security & Digital Forensics) — VIT

Jul 2018

Bhopal, India

Jul 2022

- Final grade: 1.75 Thesis: Kerberos-based authentication system

pipelines, APIs, and ML workflows.

Technical Skills

Cloud & DevOps

- Docker, Jenkins, AWS, Azure, Kubernetes (working knowledge)

Programming Languages

- Python(FastAPI, async basics), SQL, JavaScript, C#/NET, PHP, C++

Machine Learning & Data

- PyTorch, scikit-learn, time-series modelling, anomaly detection, representation learning
- Statistical analysis, model evaluation, data preprocessing

MLOps & Experimentation

- MLflow, Optuna, experiment tracking, reproducible training pipelines

Systems & Tools

- Git, Linux, PowerShell/Bash, Power BI, Power Platform, SharePoint

AI / LLM Stack

- Retrieval-Augmented Generation (RAG) pipelines
- Embeddings, vector similarity search
- LLM orchestration and prompt conditioning
- API-based integration and modular system design

Research Interests

- Time-series modelling and anomaly detection
- Representation learning and latent space analysis
- Reliable and explainable AI systems
- Industrial and energy-domain machine learning

Languages

English C2

German B1(Improving)