Nikhil Togadiya

AI/ML Engineer (Agentic AI & Generative AI)

✉ nikhil.togadiya@outlook.com
📍 Wolfenbüttel, Germany
in linkedin.com/in/nikhil-togadiya/
🔗 nikhiltogadiya.github.io/Nikhiltogadiya/
⌗ github.com/Nikhiltogadiya

🪪 PROFILE

AI/ML Engineer specializing in agentic AI and production-ready generative systems. Built autonomous multi-agent museum guides on Temi robots with RAG-based retrieval and engineered constrained LLM frameworks for autonomous vehicle test generation at Volkswagen. Hands-on expertise in LLM fine-tuning (LoRA), multi-agent orchestration, and deploying ML pipelines from research to real-world robotics and manufacturing workflows.

💼 PROFESSIONAL EXPERIENCE

09/2024 – present
Wolfenbüttel, Germany

Ostfalia Hochschule für angewandte Wissenschaften ↗
Studentische Hilfskraft (AI/Robotics) - Multi-Agent Museum Tour Guide on Temi Robots

- Architected multi-agent AI museum tour guide on 2 Temi robots with RAG-based retrieval (HuggingFace embeddings + vector DB), achieving 90%+ answer coverage with sub-1s response times via confidence-gated prompting for multilingual visitor interaction.
- Developed pytemi, a custom Python library for Temi robot control (REST API abstraction, navigation, sensors) serving as core project infrastructure.
- Mentored Master's thesis students on RAG optimization and multi-robot handover protocols, coordinating integration toward production deployment

💼 INDUSTRIAL PROJECTS

03/2025 – 07/2025
Wolfsburg, Germany

Volkswagen AG ↗
Constrained LLM Based Automated Test Case Generation for Autonomous Vehicles

- Proposed and implemented a two-stage constrained text generation framework combining formal grammar enforcement (.lark DSL) with DFA-based token masking (Syncode) to guarantee syntactic correctness of generated system test cases.
- Applied parameter-efficient fine-tuning (LoRA) on LLaMA 3B and 8B models to improve semantic alignment between requirement descriptions and generated test steps, analyzing performance trade-offs across model sizes.
- Evaluated generation quality using BLEU, SacreBLEU, and ROUGE-L against reference test cases, systematically benchmarking structural validity and semantic accuracy for autonomous driving validation tasks.

03/2024 – 07/2024
Wolfsburg, Germany

WOBCOM GmbH ↗
Energy Time Series Forecasting with RNN for Smart Metering Systems

- Developed and optimized deep learning models (RNN, LSTM, GRU) in TensorFlow for multivariate time-series forecasting of smart meter energy data; evaluated performance using MAE and RMSE and selected the best-performing architecture.
- Designed and implemented an end-to-end machine learning pipeline, including data preprocessing, feature engineering, sequence modeling, model validation, and RESTful API deployment with FastAPI for real-time inference.
- Containerized and deployed the ML application using Docker on a local server, enabling scalable prediction services and supporting data-driven energy management decisions through visualization and technical reporting.

⛶ PERSONAL PROJECTS

09/2025 – 01/2026	**Text2CAD: Multi-Agent LLM System for Autonomous 3D CAD Generation**

- Engineered an end-to-end AI pipeline that converts natural language prompts into manufacturable 3D CAD models using a 4-agent architecture (requirements analysis, planning, code generation, verification) with automated error-correction loops.
- Implemented self-learning design memory using ChromaDB vector database with RAG-based retrieval, enabling reuse of successful design patterns and iterative performance improvement.
- Integrated sandboxed CadQuery execution, Vision-Language Model–based multi-view verification, and STL/STEP export via Flask web UI and CLI for production-ready manufacturing workflows.

🎓 EDUCATION

09/2023 – present *Wolfenbüttel, Germany*	**M.Sc. Digital Technology** Ostfalia Hochschule für angewandte Wissenschaften
06/2019– 03/2022 *Surat, India*	**Bachelor in Computer Application** S.V. Patel College of Computer Science and Business Management

⚙ TECHNICAL SKILLS

Programming Languages:
Python, SQL, C++

ML & Deep Learning:
PyTorch, TensorFlow, HuggingFace Transformers, scikit-learn, XGBoost

Generative AI / LLMs:
LLaMA, OpenAI API, LoRA/PEFT, Prompt Engineering, Constrained Decoding (Syncode, DFA-based Token Masking), Formal Grammar DSLs (.lark), Vision-Language Models

Agentic AI & RAG:
LangChain, LangGraph, Tool-Use Orchestration, ChromaDB, FAISS, Qdrant Embedding Models (OpenAI, HuggingFace), Sentence Transformers, Semantic Retrieval, Confidence-Gated Prompting

NLP:
Text Generation, Semantic Similarity, Multilingual NLP, Semantic Search, Tokenization, Evaluation (BLEU, SacreBLEU, ROUGE-L)

Deployment & Tools:
Docker, FastAPI, Flask, Ollama, REST APIs, Git, Linux, Jupyter

Data & Libraries:
Pandas, NumPy, Matplotlib, PostgreSQL, Feature Engineering

Computer Vision:
OpenCV, YOLO, Multi-View Verification

🧠 SOFT SKILLS

- Team Player
- Quick Learner
- Adaptable
- Flexible
- Good Communication
- Active Listener

🌐 LANGUAGE

- English - Fluent
- German - B1 Level