Yu-Sheng (Adam) Tang

Generative AI Engineer, Data Scientist | 5+ Years of Expertise, Freiburg im Breisgau, Germany
+49 1578-129-1537 | adam951502@gmail.com | adam951502 | yu-sheng-tang | adam951502.github.io

Education

University of Freiburg **Germany** **Oct 2018 - Feb 2021**
Master of Science in Sustainable Systems Engineering German Grade: 1.7 (1.0 = best), GPA: 3.6/4.0

Chang Gung University **Taiwan** **Sep 2012 - Jun 2016**
Bachelor of Science in Mechanical Engineering German Grade: 1.3 (1.0 = best), GPA: 3.8/4.0
* Dean's list Award: Feb 2014, Sep 2014, Feb 2015

Skills

Languages: Python, MATLAB, C++, Bash, LaTeX
DBMS: PostgreSQL, MySQL, GraphDB (SPARQL), Neo4j
Web: HTML/CSS, JavaScript, Bootstrap, Flask, RESTful APIs, Swagger, JWT
Data Skills: ETL, data wrangling, web scraper, web crawling, data warehousing, data catalog
Tools & Techs: pypi, pip, git, pgadmin4, International Data Space
DevOps: CI/CD (GitHub, GitLab), AWS, Jenkins, Pytest, Docker, uWSGI, NGINX
LLMs & RAG: Ollama(Llama3, DeepSeek), Open WebUI, LangChain, Chainlit
Management: Jira & Confluence (admin, maintainer), project management, agile management, requirement engineering
Certificate: IBM Data Science Professional Certificate, International Requirement Engineering

Experience

Generative AI Engineer | Fraunhofer EMI | Germany **Oct 2021 - Present**
Technical Skills:

- Designed an ontology for over 3,000 datasets and built a knowledge graph (KG) for material life cycle assessment (LCA) of additive manufacturing processes using Python and SPARQL, enhancing data pipeline, analysis, and query capabilities by 50%.
- Designed architecture and use-case of resilience data space.
- Constructed an efficient ETL data pipeline in Python, significantly enhancing efficiency by 80%.
- Established a Python package and RESTful APIs (Flask) to streamline data processing, enhancing productivity and code reusability by 90%.
- Established microservices backend APIs using Flask, uWSGI, NGINX, and Docker, reducing deployment time by 80%.
- Implemented DevOps and CI/CD pipeline for automated testing, release, and deployment, enhancing code quality and project efficiency in GitLab by 40%.
- Established a complete ETL pipeline for the public health database using PostgreSQL, pgAdmin4, GraphDB, Flask, uWSGI, NGINX, and Docker, improving data accessibility by 80%.
- Deployed and integrated large language models (LLMs), Llama, using Ollama, Open WebUI, Chainlit, and Docker to enhance the (meta)data pipeline, increasing processing speed by 90%.
- Developed a RAG architecture application using Ollama, Chainlit, and Docker to reduce LLM hallucinations by integrating a knowledge base with LLM-generated responses for greater accuracy and reliability.

Project Management:

- Supervised a master's student, providing guidance throughout the course of the master project (with German Grade: 1.0) – *Designing An Automated Metadata Extraction Pipeline: Bridging the Gap Between Data Collection and Unified Graph-Based Metadata Representation*, resulting in an 85% completion rate.
- Supervised 2 students in establishing an ETL pipeline for a comprehensive metadata graph that describes heterogeneous data from various sources, achieving a 90% increase in data processing efficiency.
- Developed a universal use-case guideline and template, significantly enhancing the efficiency and quality of software development processes within the HERAKLION project, leading to a 30% improvement in project delivery times.
- Managed requirements (requirement engineering), agile project management cycle, and project execution for a team of more than 20 people using Jira, streamlining workflows and improving efficiency by 35%.

Projects

FEM Simulation Assistant Chatbot – Local RAG-Powered Recommender System for Automotive R&D **May 2024 - Present**
Tech Stack: LLM(Ollama-Llama3, DeepSeek), Chainlit, RAG, Docker, Python, Database

- Developed a Retrieval-Augmented Generation (RAG) system to minimize LLM hallucinations by integrating document retrieval from a domain-specific knowledge base with AI-generated responses, ensuring higher accuracy, contextual relevance, and reliability in outputs.

HERAKLION - *Heuristic Resilience Analyses for Municipalities Using Data Space Functionalities* **Mar 2022 - Present**
Tech Stack: pip, git, CI/CD, DevOps, ETL, RESTFul API, Python, py-library, Flask, Docker, KG, requirement engineering, agile management, RAG

- Developed a scalable resilience data space demonstrator to improve crisis data accessibility and usability for German municipalities and emergency services, facilitating timely and informed decision-making.

ADAM-SusTrace - *Networking of digital assets and data-driven value creation through data ecosystems in additive manufacturing* **Oct 2021 - Feb 2022**
Tech Stack: git, Python, algorithm, data structure, Ontology, knowledge graph, LCA

- Applied digital traceability in additive manufacturing to conduct comprehensive sustainability analyses (Life Cycle Assessment, LCA). Assisted in developing tools and workflows, enhancing data ecosystem's value, and facilitating detailed understanding of product/process sustainability.