Zi Qiao Dai

Software Engineer | Backend & Applied AI
ziqiaowx.dai@gmail.com | Singapore Citizen | LinkedIn

Summary

Software engineer focused on backend and applied-AI systems, with experience building AWS-hosted FastAPI applications, Jira ingestion pipelines, retrieval tools, and multi-agent workflows at Deloitte. Experienced in PostgreSQL/pgvector, sentence-embedding training, hybrid search, and reproducible NLP evaluation.

Skills

Languages: Python, TypeScript, Rust
Backend & Data: FastAPI, REST APIs, Python asyncio, SQL, PostgreSQL, pgvector, Jira API/JQL
AI/ML: PyTorch, Sentence Transformers, Hugging Face Transformers, RAG, LLM agents and tool calling, contrastive learning, BM25 and hybrid retrieval
Platforms & Tools: AWS Bedrock, Claude, Cohere Embed/Rerank, Model Context Protocol (MCP), Git, Playwright

Experience

Analyst *(Software & AI Engineering)* Sep 2025 – Present
Deloitte – Technology & Transformation, Singapore

- Built a multi-agent UAT-generation feature in **FastAPI** using **Claude via AWS Bedrock**, reducing analyst-estimated initial drafting time for a complex end-to-end UAT script from **approximately 3 person-days to 10–15 minutes**; generated **50+ scripts**, most requiring only minor formatting edits.
- Developed an asynchronous Jira ingestion and retrieval pipeline using **Cohere Embed**/**Rerank**, **pgvector**, and a JQL fallback to surface related tickets and epics for agent workflows.
- Shipped a citation-backed RAG assistant adopted by **three internal teams**, grounding project-specific answers in original Jira tickets.
- Built an agentic test-automation workflow integrating Jira and **Playwright MCP** to retrieve requirements, inspect web interfaces, generate and execute Page Object Model-based TypeScript tests, and iteratively repair failures before human review.

Selected Projects

Sentence-BERT Training & Evaluation 2026
Nanyang Technological University

- Trained Sentence-BERT models in **PyTorch** using NLI warm-start, contrastive and ranking objectives, and STS fine-tuning; compared objective and attention-pooling variants on a fixed `bert-base-uncased` backbone.
- Across three seeds on STSBenchmark, the best MNRL+UniReg model achieved a **cosine-Spearman score of 86.8**, **1.9 points above** the plain STS baseline without changing the encoder architecture.

Hybrid Retrieval for Multi-Hop Question Answering 2026
Nanyang Technological University

- Implemented a **hybrid retriever** that normalized and fused **BM25** lexical scores with bge-base-en-v1.5 dense similarities, balancing exact entity matching with semantic retrieval of multi-hop bridge evidence.
- Compared BM25 and hybrid retrieval under identical cross-encoder reranking on a 200-question HotpotQA sample; hybrid increased **Recall@10 from 0.950 to 0.963**, while **MRR@10 remained 0.962**.

Education

Master of Science in Artificial Intelligence Jan 2026 – Jun 2027 (Expected)
Nanyang Technological University, Singapore

Bachelor of Computer Science Feb 2023 – Sep 2025
Monash University, Melbourne