

ShijieZheng-AI Engineer(Master's Degree)

+86 18772198735
z1597006376@gmail.com
<https://www.linkedin.com/in/shijie-zheng-8b230433b/>
<https://github.com/Formyselfonly>

Education

Hubei Normal University-Electronic Information Engineering	September 2018 - June 2022
Yangtze University-Master's Degree (Artificial Intelligence)	September 2022 - June 2025
<ul style="list-style-type: none">Research Focus: NLP (Natural Language Processing)SCI Paper: MemoryRepository for AI NPCNational Award: Awarded the 2024 China Graduate National Scholarship	

Professional Skills

- Knowledge:** Proficient in core technologies and methods in **Machine Learning (ML)**, **Natural Language Processing (NLP)**, and **Generative AI**. Familiar with related models and frameworks, with extensive project experience.
- Prompt Engineering:** Skilled in designing and applying **formatted Prompts**, **Few-shot Learning**, **Prompt Reverse Engineering**, **Role-playing Prompts**, and **Tool-based Prompts**. Capable of **automated Prompt evaluation** and **directed optimization iteration**.
- LLMs:** Proficient in the **development, training, and integration** of mainstream large models like **ChatGPT**, **Llama3**, **DeepSeek**, **Gemini**, and **Claude**, with experience in multiple production projects.
- API Development:** Expert in using **Python**, **Flask**, and **FastAPI** to develop and deploy APIs. Familiar with encapsulating model services, optimizing requests, and integration, with practical engineering experience.
- Large Model Frameworks:** Proficient in using **Dify**, **Coze**, and **RAGFlow** to build **AI project demos**, and familiar with frameworks such as **LangChain**, **LangGraph**, and **LangSmith** to build **advanced AI applications**.
- AI Agent:** Skilled in using **RAG technologies**, **Workflows**, and **Tools** to build **AI Agents**, with expertise in **RAG document processing**, **Chunking**, **parameter testing**, and optimization. Able to abstract tasks into **Workflows** and gradually improve them.
- Project Deployment:** Familiar with **Linux systems** and proficient in using **Docker** for deploying common **AI projects**, such as **Ollama+OpenWebUI**, **Dify+RagFlow**, **ComfyUI**, **Chatbox**, and **AutoGPT**.
- Cutting-edge Research:** Continuously exploring emerging technologies. Familiar with **MCP (Multi-Agent Collaboration)** and **DeepSearch**. **DeepSearch** is a self-developed solution inspired by the design of **OpenAI's DeepResearch**, independently implemented through **LLM**, **Tavily**, and customized **Workflow orchestration**.
- AI Solutions:** With **three years** of experience in generative **AI development**, familiar with the technical landscapes and solutions in the fields of **e-commerce**, **gaming**, and **efficiency tools**. Offers personal insights into the **AI field** and can propose practical AI implementation solutions.

Internship Experience - OgCloud (Guangdong Tianyun Technology Co., Ltd.)

Duration: 2023.04 - 2024.04
Position: AI Intern

OgAI

Project Description: Developed internal AI-powered tools to enhance company efficiency.
Responsibilities: **AI solution design** + **Prompt engineering** + **Python backend development**

AI Game Companion

Project Description: AIGC-based game platform.
Responsibilities: **Game prompt design** and **model deployment and fine-tuning**, building **multi-agent architecture** and implementing demo

XiaoO Assistant Bot - AI Enterprise Assistant + AI Intelligent Customer Service

Project Description: XiaoO is a company-customized assistant built with **RAG** technology based on internal documentation, and supports **system-level command execution**.
Responsibilities: Built with **Dify** framework, implemented a demo of **OsGPT** for system-level commands, conducted **chunking and retrieval testing** for the RAG document base

MemoryRepository System for Long-Dialogue NPC Agents

Project Description: Humanized memory system for LLMs in role-playing applications
Responsibilities: Solving the problem of context limitations in LLM dialogues through algorithm mechanisms, resulting in the publication of the *SCI paper MemoryRepository for AI NPC*

Internship Experience - Ximalaya (Shanghai Ximalaya Technology Co., Ltd.)

Duration: 2024.05 - 2024.06
Position: AI Intern

AI Script Generation

Project Description: Utilized **RAG** and **workflow orchestration** to automate generation, expansion, adaptation, evaluation, and iteration of AI-generated scripts from a novel database
Responsibilities: Responsible for **prompt design, evaluation, and optimization** in the script generation workflow

Internship Experience - The Home Depot (Shanghai)

Duration: 2024.10 - 2025.02
Position: AI Intern

AI Document Processing Tool

Project Description: Built intelligent agents using **RAG** and **workflow** to classify and summarize internal reports and assist analysis with data insights
Responsibilities: Deployed local models (**Ollama + OpenWebUI**), built **RAG + Workflow** agents to implement the AI document processing tool demo

Project Experience

Personal Project Repository: <https://github.com/Formmyselfonly/>
(These are online demo-level projects for display,main personal projects will be introduced during interviews)
Main personal projects: **MultiAgent**, **OsGPT**, **MemoryRepository**, **AI Search**, **Prompt Eval**

ECommerceGPT - AI E-commerce Efficiency Website

<https://github.com/Formmyselfonly/ECommerceGPT>
Developed an AI efficiency tool for the eCommerce sector, leveraging ChatBox and prompt engineering. Comprehensive usage documentation is available on GitBook, enabling users to select from various templates tailored for enhancing ECommerce efficiency.

PromptEval - Prompt Response Analyse

<https://github.com/Formmyselfonly/PromptEval>
Prompt Evaluation Approach: Combining Data Evaluation and Human Evaluation
Multidimensional Evaluation: Evaluating the quality of Prompt Responses through seven dimensions..
Human Evaluation: The project includes automated scripts for batch testing of prompts, followed by manual evaluation by human trainers.

AI Project Example in Dify - Companion Dialogue AI Agent Robot

Leveraging LLM, prompt engineering, and RAG, this project provides an API and an online demo.
Ganyu from Genshin Impact - <https://udify.app/chat/iP0yw8HbPFed6XzL>
Honkai Impact Simulator - <https://udify.app/chat/Jf7UHHnVI8mpZb9G>

Certificates & Language

- English *CET-6*
- Chinese *Native*
- SCI Paper *MemoryRepository for AI NPC*