

Galen Wu

IDP Backend Developer
18584883573 - wuguanlin312@gmail.com
Chengdu, China
Date of birth 1994.10

Profile

- Highly skilled AI Algorithm Engineer with extensive experience in **Python development, RESTful APIs (FastAPI)**, and robust database handling.
- Proficient in **Computer Vision, Image Processing**, and **OCR technologies** for **document processing**.
- Proven ability in **Multimodal** Generative AI (GPT, Qwen-vl-Max) and fine-tuning pre-trained models using **Hugging Face, OpenAI, PyTorch/TensorFlow**.
- Experienced in deploying Python applications via **Docker**
- Strong advocate for **Git** Version Control, code reviews, and advanced Git operations.
- A collaborative team player with excellent communication skills, adept at prioritizing tasks, debugging, and continuously learning new **AI/document processing tools and frameworks**.

Education

Sichuan University of Science and Engineering | 2013/9 – 2017/6

Employment History

Algorithm Engineer, Chengdu Huineng Boyou Technology Co., Ltd.

March 2024 – May 2025

- Focused on AI product R&D, AGI architecture, and algorithm development.
- Lead 5-10 member teams

Achievements:

1. Fourteen enterprises have been approved for the project
2. Enhance the core competitiveness of the product

Senior Python Development Engineer, ThunderSoft (Chengdu) Co., Ltd.

July 2023 – January 2024

- Led core R&D for strategic products, participating in requirements analysis and POC phases.

Development Technical Expert, DataGrand (Chengdu) Co., Ltd.

June 2020 – July 2023

- Managed software development projects

Achievements:

1. Recognition as the sole search expert in the Southwest region
2. Two A-grade promotions.

Python R&D Engineer, Chengdu Youxinli Intelligent Technology Co., Ltd.

January 2019 – May 2020

- Responsible for MES project construction and management.

Python Development Engineer, Chengdu Tianfu Human Resources Development Service Co., Ltd. December 2017 – December 2018

- Developed and maintained software modules, including coding and testing.

Projects Description

Project: LLM-based Q&A System
Duration: July 2023 – Present

Overview:

A large language model-powered question answering system that integrates retrieval-augmented generation (RAG), document parsing, multimodal input handling, and intelligent intent rewriting to deliver accurate and responsive answers in both general and FAQ scenarios.

My Responsibilities:

- Led development and core algorithm design, focusing on domain-specific LLM enhancements.
- Built and optimized intent rewriting models using Transformer-based architectures.
- Solved document parsing and image-based Q&A challenges via multimodal techniques.
- Designed core components including chunking, retrieval, re-ranking, prompt templating, and streaming media nodes.
- Applied Langchain principles to implement chain-of-thought reasoning and multi-path recall.

Achievements:

- Achieved 85%+ accuracy on general QA and 97% on FAQs, with 92% recall coverage.
 - Improved precision recall via two-stage intent rewriting and disambiguation techniques.
 - Successfully deployed a full-stack LLM Q&A system across multiple business use cases.
-

Project: Knowledge Graph Q&A Platform

Duration: July 2022 – July 2023

Overview:

A Q&A system powered by a knowledge graph, enabling users to query structured domain knowledge with high accuracy and context-awareness. The system integrated document parsing, semantic search, and recommendation algorithms to enhance user interaction and answer relevance.

My Responsibilities:

- Led internal requirement analysis and guided clients in defining their core needs.
- Served as chief engineer, designing the development framework and overseeing API integration.
- Applied advanced techniques including search ranking, knowledge graph algorithms, watermarking, and document parsing (Word/Excel/PDF).
- Estimated workloads, managed schedules, ensured quality delivery, and maintained post-launch stability.

Achievements:

- Delivered a stable, scalable knowledge graph Q&A system through multiple product iterations.
 - Successfully met all project milestones with high delivery quality and long-term system reliability.
-

Project: Intelligent Stability Maintenance System

Duration: July 2022 – July 2023

Overview:

An intelligent early warning and behavior prediction platform designed to support social stability efforts. The system integrated real-time data synchronization, risk modeling, and alert mechanisms to proactively identify and assess potential stability-related events.

My Responsibilities:

- Designed system architecture and selected databases; implemented MySQL-binlog-based real-time data synchronization.
- Developed core modules including risk scoring, judgment models, and behavior prediction using AHP, tanh functions, and priority logic.
- Built and refactored key workflows (e.g., clue and personnel reporting); prepared test datasets and implemented unit tests.
- Solved technical challenges (e.g., Elasticsearch7 & Mongoengine integration), refined algorithms, and optimized system architecture.
- Led code quality efforts using Pytest, FactoryBoy, Celery, and pre-commit checks.

Achievements:

- Delivered a reliable and intelligent risk monitoring system with proactive early warning capabilities.
- Significantly improved stability event detection accuracy and system robustness through algorithm optimization and architectural refinements.

Project Group: AI Productions

Duration: June 2020 – July 2023

1. Knowledge Graph Big Data Project Overview:

A high-throughput knowledge graph platform built on Neo4j and Hive to process and ingest billion-scale data with automated relationship mapping.

My Responsibilities:

- Led Neo4j adaptation for large-scale data and handled full-cycle delivery.
- Designed and deployed automated data pipelines using Hive and big data tools.
- Navigated ambiguous requirements through multiple design iterations to ensure high delivery quality.

Achievements:

- Successfully delivered a scalable knowledge graph ingestion pipeline for billion-level datasets.
- Reduced manual mapping workload through relationship automation.

2. Intelligent Q&A Platform Overview:

A semantic search and question answering system integrating long-text retrieval and Elasticsearch optimization.

My Responsibilities:

- Co-developed long-text segmentation algorithms with the algorithm team to enhance Elasticsearch Q&A accuracy.
- Finalized deliverables in collaboration with PMs, testers, and HQ, including DB design documentation.

Achievements:

- Improved recall accuracy in Q&A tasks through custom text segmentation.
- Ensured delivery readiness with complete documentation and testing alignment.

3. Search Product – Mid-end Project Overview:

A reusable search infrastructure project to streamline internal product integration across regions.

My Responsibilities:

- Acted as regional lead for Southwest projects, delivering plug-and-play search product solutions.
- Created and maintained technical documentation and led training to reduce repeated mistakes.

Achievements:

- Enabled efficient adoption of search products across multiple teams.
- Minimized recurrent errors through proactive documentation and knowledge sharing.

Cross-Project Responsibilities:

- **Leadership:** Chief Engineer and Regional Head, coordinating across teams and projects.
 - **Technical Expertise:** Proficient in Elasticsearch, Neo4j, Kafka, Hive, and large-scale data processing.
 - **Problem Solving:** Specialized in scalability, algorithm tuning, and cross-functional troubleshooting.
 - **Knowledge Sharing:** Strong focus on documentation and reusable technical assets.
-