Gaosong Chen

+49 152 92609019 cedarchen7@gmail.com Berlin, Germany
https://www.linkedin.com/in/cedarchen https://tallsong.github.io
Software Engineer Intern

SUMMARY

Master's student with 5 years of experience in software development, specializing in AI agent development. Proven ability to build and deploy AI agents, demonstrated by the creation of a mental health therapy agent that improved counseling session length and relationship ratings. Possesses expertise in Large Language Model (LLM) performance optimization, including designing and implementing LLM inference API services.

EDUCATION

XU Exponential University of Applied Sciences 2025.04 - 2027.03
Industry 4.0 Master Potsdam, Germany

Guangdong Baiyun University 2016.09 - 2020.07
Information Management and Information System Bachelor Guangzhou, China

SKILLS LIST

- Programming Languages: Python, C, C++, JavaScript, Shell
- Frameworks/Tools: PySpark, PyTorch, CUDA, Django, Qt, Flask, Celery, MySQL, Redis, Git, Docker, Kubernetes, Linux, Scrapy, CEPH, RESTful API, Cursor, Gemini, Grok, LangChain
- Testing Platform: Ansible, Pytest
- Language: English(Professional), Chinese(Native Speaker), German(Conversational)

RESEARCH EXPERIENCE

Mental health therapy agent 2025.06 - Present
Research Intern @Tsinghua University remote

- Implemented code for AI agent(Python,LangChain, ChatGPT4o, Vue.js)
- Integrated a hierarchical three-layer planning system with bidirectional adaptation, orchestrates communication through parallel guidance and empathy pathways, and manages dynamic, context-aware interventions via comprehensive client profiles and a structured knowledge base
- **Achieved a 95.4% increase** in average session length and a 46.7% improvement in overall counseling relationship ratings

PROFESSIONAL EXPERIENCE

Runjian Co. 2023.02 - 2025.04
Software Engineer Beijing, China

- Led the development of word processing software that automatically identifies Word documents, extracts content, and renders it in a GUI for user interaction, **increasing UI response efficiency by 95%** (Qt, C++)
- Designed and implemented a Python/Flask asynchronous inference API integrating the VLLM engine, 4-bit quantization (bitsandbytes), CUDA-accelerated PyTorch, and Docker containerization, which achieved a **2× increase in throughput** and a **70% reduction in memory usage**, dramatically improving response speed

JD.com 2021.10 - 2022.08
Software Development Engineer Beijing, China

Participated in the development of Nine Security Computing Platform, mainly responsible:

- SQL Parser: Generate AST by parsing the SQL input by the user through Lexer and parser, and then perform code generation to generate the corresponding secure computing API code (ANTLR, Pyspark, SQL, Python), and the **supported SQL syntax coverage increased by 50%**
- Jupyter development: developed custom user login, code detection, custom kernel, and other functions (Jupyter, Python, Kubernetes, Docker)
- File service: developed user data storage, query, and other related functions (Flask, Pandas), supports access to file list of folders from the file system, HDFS, and S3 interfaces, creating folders, obtaining CSV and ORC format headers, obtaining the number of rows and columns, file size, uploading files, and reading the first N lines of files. The results are returned in restful style JSON, with the average response time reduced to 200ms

Institute of Information Engineering, Chinese Academy of Sciences 2020.07 - 2021.10
Software Development Engineer Beijing, China

- Optimized the user interface and implemented the backend functions (C++, QT, socket) of the security and security authorization management system of the virtualization platform.
- Designed and implemented a distributed storage system, mainly implementing user management, uploading and downloading files, file collection and sharing, user history viewing, and other functions (CEPH, Linux, Django, Boto3, HTML, CSS, JavaScript, python)
- Storage cluster monitoring module development and document online browsing function implementation (Flask, JavaScript, Python)

Yuexin Technology 2019.06 - 2019.12
Data Development Engineer Intern - Big Data Engineering Beijing, China

- Extracted, scheduled, processed, cleaned, and analyzed hundreds of millions of data levels in the data warehouse.
- Utilized ETL tools to write scripts and SQL for data processing, **handling approximately 4TB of data daily** (Kettle, JavaScript).
- Developed data crawlers for enterprise industrial and commercial data (Python, Scrapy, XPath).