

Gioia Zheng

Rome, Italy | zheng.2040439@studenti.uniroma1.it | linkedin.com/in/gioia-zheng-9233a0303
github.com/GioiaZheng | gioiazheng.github.io | ORCID: 0009-0000-8212-4506

Profile

Applied Computer Science and AI undergraduate at Sapienza University of Rome, focused on information retrieval, retrieval-augmented generation, evaluation methodology, and backend systems. I build reproducible research-engineering pipelines: retrieve/rerank/generate workflows, paired statistical evaluation, experiment manifests, CI checks, and inspectable documentation. Open to research internships, ML/research-engineering roles, and backend/software engineering roles in Italy, the EU, or remote.

Education

Sapienza University of Rome

Sep 2022 - Present

B.Sc. in Applied Computer Science and Artificial Intelligence

Research and Engineering Experience

NLP Research Intern, Microsoft

Feb 2026 - May 2026

Natural Language Computing, remote

- Designed and implemented a multi-stage retrieval and generation evaluation pipeline over MS MARCO-scale data: BM25 retrieval, FAISS dense retrieval, transformer reranking, and seq2seq generation.
- Built statistical evaluation routines for per-stage lift, including semantic and exact-match generation metrics, paired comparisons, and confidence intervals.
- Engineered CI and reproducibility support for long-running indexing and evaluation jobs, including report-ready outputs and experiment notes.

Web Development Intern, Idearia

Mar 2026 - Present

Rome, Italy

- Develop and maintain PHP / WordPress backend systems for client-facing production websites.
- Build n8n automation workflows connecting WordPress forms, content pipelines, and external services, reducing manual operational work.

Selected Projects

MS MARCO Generative QA System

github.com/GioiaZheng/msmarco-genqa

Python, PyTorch, BM25, FAISS, cross-encoder reranking, T5, paired bootstrap

- Built a reproducible retrieve -> rerank -> generate pipeline on MS MARCO passage data, with config-driven experiments and schema-versioned manifests.
- Reported Token-F1 improvement from 0.197 to 0.368 on 6,980 paired queries, with paired-bootstrap confidence interval [+0.163, +0.178].
- Documented limitations clearly: dense/reranker experiments use a qrels-anchored sample; CPU-only reproducibility is prioritized over GPU throughput.

Handwritten OCR Pipeline

github.com/GioiaZheng/handwritten-ocr-system

PyTorch, CNN, BiLSTM, CTC, IAM handwriting dataset

- Implemented a handwritten text recognition pipeline from preprocessing to CRNN/BiLSTM/CTC decoding, with CER/WER evaluation hooks.
- Added reproducibility documentation for training configuration, seed, deterministic split, and pending measured CER/WER reporting.

Go Chat Service

github.com/GioiaZheng/go-chat-system

Go, Vue, SQLite, OpenAPI, Docker Compose

- Built a coursework-grade full-stack chat service with REST endpoints, SQLite persistence, Vue UI, OpenAPI documentation, API tests, and a documented authentication threat model.
- Kept production boundaries explicit: simplified token auth, single-node SQLite deployment, and no production security claims.

CiboCompass

github.com/GioiaZheng/CiboCompass

React Native, Expo, Go REST API, SQLite, offline-first UX

- Built a mobile app prototype for international students reading restaurant menus, including dish lookup, cultural rating comparison, and local rating persistence.
- Documented offline sync states and demo assets for review without overstating production readiness.

Additional Experience

Young Representative, UPI Emilia-Romagna

Sep 2022 - Oct 2023

Forlì-Cesena province

- Conducted data-driven research on school dropout and contributed to policy-oriented reports.

Programming Intern, Sandbox Games

May 2021 - Jun 2021

Emilia-Romagna, Italy

- Contributed to a 3D simulation platform, working on backend systems and MySQL query optimization.

Awards and Certifications

- 1st Place, "WISWUP" Business Plan, A Scuola di Impresa Competition.
- TryHackMe Pre-Security Learning Path Certificate; Cyber Security 101 Certificate.

Skills

Programming

Python, Go, SQL, JavaScript, PHP

ML and Retrieval

Information retrieval, RAG evaluation, PyTorch, FAISS, BM25, Sentence Transformers, cross-encoder reranking, T5

Evaluation

Paired bootstrap, confidence intervals, Token-F1, ROUGE-L, BLEU, exact match, CER/WER

Backend and Web

REST APIs, OpenAPI, SQLite, Vue, WordPress, n8n automation, Docker, GitHub Actions

Languages

Chinese (Native), Italian (Bilingual), English (Professional), German (Beginner)