Gioia Zheng

B.Sc. Applied Computer Science and Artificial Intelligence, Sapienza University of Rome

RESEARCH INTERESTS
gioiazheng.github.io | GitHub
LinkedIn | gioia.zheng.stud@gmail.com

I study how retrieval-augmented systems fail, how those failures change across configurations and data boundaries, and how evaluation artifacts can make the evidence inspectable and reproducible.

Research focus

01

Evaluation beyond one aggregate score

I am interested in paired comparisons, uncertainty estimates, controlled ablations, and per-query analysis. The aim is to distinguish a robust improvement from a result that depends on one metric, threshold, or sample.

02

Retrieval and evidence diagnostics

I want to separate retrieval miss, ranking error, context pollution, evidence omission, unsupported generation, and citation failure. This requires traces that preserve intermediate states rather than only the final answer.

03

Reproducible ML systems

I treat schemas, experiment manifests, configuration checks, tests, and explicit limitations as part of the research result. A claim should remain traceable to the code, data, configuration, and evaluator that produced it.

Evidence already built

RAG Observatory

Trace schema, OpenTelemetry/OpenInference JSON ingestion, stage-aware reports, configuration comparison, 40 public SciFact retrieval traces, and 107 automated tests.

CiboCompass

A React Native/Expo client with a Go REST API and SQLite persistence. The system documents its offline-delivery guarantee, retry behavior, idempotent writes, and limits around multi-device synchronization.

MTEB contributions

Two merged upstream changes: evaluation-request documentation and model language-scope metadata with tests and accessibility coverage.

Next question and research fit

PLANNED NEXT STUDY

Measure chunk-boundary and segmentation sensitivity on public SciFact evidence while holding retrieval and evaluation settings fixed. The goal is to identify when evidence coverage changes only because a supporting sentence crosses a chunk boundary.

OPPORTUNITIES SOUGHT

M.Sc., HiWi, research assistant, and research engineering opportunities in information retrieval, NLP, reliable RAG, and reproducible ML systems, particularly in Germany and Europe.

Methods I use

CONTROL

Fixed query sets, one named variable, and invariant configuration checks.

UNCERTAINTY

Paired evaluation, bootstrap intervals, and exact counts alongside means.

ATTRIBUTION

Stage-local signals tied back to retrieval, context, generation, or evaluation evidence.

REPRODUCTION

Versioned schemas and manifests, automated tests, CI, and documented scope limits.