Francesco **Congiu**

PhD Student

francesco.congiu@acm.org | www.francesco.congiu.com | wakaflocka17 | f-congiu-res

Internship

Apply
Cagliari , Italy

Applied Scientist — *Mar 2025 - Jul 2025*

- Engineered a Confidentiality-Preserving RAG system using Python and HuggingFace, ensuring data privacy for enterprise-level document retrieval.
- Optimized retrieval accuracy by integrating custom EM-based reranking algorithms, reducing false-positive rates in query responses.
- Collaborated with the engineering team to deploy scalable LLM pipelines, improving internal document processing speed.

Adiacent
Cagliari , Italy

Full Stack Developer — *Mar 2023 - Jul 2023*

- Architected a document digitization pipeline using C# and .NET, automating the ingestion of legacy records into modern databases.
- Designed a core infrastructure module currently utilized to manage 5.000+ digital assets.

Education

University of Cagliari (UniCa)
Cagliari, Italy

Ph.D. in **Teaching & Learning Sciences — Specialization in AI & Educational Technologies** — *Nov 2025 – Current*

- **Research Areas:** Retrieval-Augmented Generation Systems, Confidentiality-Aware Systems, Natural Language Processing, Information Retrieval.

University of Cagliari (UniCa)
Cagliari, Italy

M.S. in **Applied Artificial Intelligence**, SUMMA CUM LAUDE — *Sep 2023 - Sep 2025*

- **Thesis**: "RetrievEM: Confidentiality-Preserving RAG via Expectation-Maximization"
- **Relevant Coursework**: Information Retrieval, Natural Language Processing, Machine Learning, Deep Learning.

University of Cagliari (UniCa)
Cagliari, Italy

B.S. in **Computer Science** — *Sep 2020 - Jul 2023*

- **Thesis**: "Development of a web application for blood donation management".
- **Relevant Coursework**: Algorithms & Data Structures, Software Engineering, Programming 1, Programming 2, Operating Systems

Publications

Confidential Retrieval-Augmented Generation in Educational Contexts
Cagliari, Italy

WAILS 2025 — To appear in Springer LNCS Post-Proceedings — *Oct. 2025*

- **Authors:** Boratto L., **Congiu F.**, Fenu G., Medda G., Pau A.
- Validated confidentiality-preserving architectures by benchmarking against Beir/FiQA dataset, achieving a MAP@10 of 0.3506 through an optimized Linear Fusion strategy.
- Introduced *Backfill*, a search-depth expansion algorithm improving HIT@10 by 0.09, balancing pedagogical utility with strict access-control policies.

Research Projects

RetrievEM: Open-source Framework for Confidentiality-Aware RAG
GitHub

Creator & Lead Maintainer — *Sep. 2025*

- Designed and open-sourced a modular RAG framework integrating dense retrieval, cross-encoder reranking, and score-level fusion.
- Implemented an Expectation-Maximization (EM) inspired reranking algorithm to maximize query relevance while enforcing strict 100% confidentiality constraints.
- Developed a synthetic persona generation module to enable access-aware evaluations on benchmarks lacking native confidentiality metadata.

Key Skills

Languages	Python, SQL, C++, C#, Java, LaTeX
AI & Machine Learning	PyTorch, HuggingFace, RAG, LangChain, EM Algorithm, NLP, IR
Tools & Infrastructure	Docker, Git, Linux (Ubuntu/Arch), .NET, MongoDB, PostgreSQL, Qdrant