

Federico Borazio

Rome, Italy · borazio@ing.uniroma2.it · +39 3924532216

LinkedIn: federico-borazio · Google Scholar: Federico Borazio · Portfolio: fedbor.github.io

PROFESSIONAL SUMMARY

Data Scientist and Applied Scientist specializing in AI, NLP and Large Language Models. Ph.D. candidate with experience in developing open-source frameworks (CALAMITA) and deploying AI solutions for high-stakes domains (Healthcare, Nuclear Safety). Skilled in bridging the gap between theoretical research and industrial application, with a strong track record of publications in top-tier venues (ACL, ECIR).

EDUCATION

University of Rome “Tor Vergata” Rome, Italy
Ph.D. Candidate in Data Science 2023 – Present

- Research focus: Sustainable LLM adaptation and evaluation, RAG systems, Biomedical Question Answering systems.

University of Rome “Tor Vergata” Rome, Italy
Master’s Degree in Computer Science (Grade: 110/110 cum laude) Obtained 2023

EXPERIENCE

Amazon.com Los Angeles, United States of America
Applied Scientist II Mar 2026 – Present

- **Alexa Local Search:** Working on **LLM** training and evaluation on **Neural Retrieval** for local search and Point-of-Interest Ranking. Data Analysis and automatic **Data Annotation** pipelines.

Reveal S.r.l. Rome, Italy
AI Project Consultant & Engineer Jan 2024 – Mar 2026

- **Industrial Matchmaking Platform:** Engineered the backend for a “Match-Making for Innovation” platform (Rome Technopole), utilizing **Python**, **SentenceBERT** for semantic indexing, and **Apache Solr** for retrieval.

University of Rome “Tor Vergata” – Semantic Analytics Group Rome, Italy
AI Researcher & Developer Nov 2023 – Present

- **Project CALAMITA (Open Source):** Researcher & Developer for the first unified evaluation framework for Italian LLMs. Contributed to the implementation of several tasks to benchmark generative models within a nationwide community initiative.
- **Epidemic Intelligence (with ISS):** Developed LLM-based pipelines for the Italian National Institute of Health to detect emerging epidemiological threats from online data streams.
- **Nuclear Safety (with IAEA):** Engineered a semantic search engine and document management system to enhance transparency in nuclear decommissioning processes.
- **BioASQ Challenge Participant:** Developed *UniTor*, a modular RAG system for biomedical QA. Combined dense retrieval and generative models, achieving strong performance in the international BioASQ competition.

TECHNICAL SKILLS

- **AI & NLP:** Large Language Models, RAG, Transformers, PyTorch, Semantic Search.
- **Programming:** Python (Advanced), Java, SQL.
- **Tools & Platforms:** Git, Apache Solr, PostgreSQL, LaTeX, Generative AI tools (Cursor, Kiro, Amazon Q).
- **Languages:** Italian (Native), English (Full Professional Proficiency, Duolingo Score: 120/160).

SELECTED PUBLICATIONS

Author of 12+ peer-reviewed papers in international conferences (EMNLP, ACL, ECIR)

- **ACL 2025:** “Training Multi-Modal LLMs through Dialogue Planning for HRI” (Findings).
- **ECIR 2025:** “Adapting llms for domain-specific retrieval: A case study in nuclear safety”.
- **ECIR 2026:** “Integrating AI and IR Paradigms for Sustainable and Trustworthy Accurate Access to Large Scale Biomedical Information”.

AWARDS & OTHER CONTRIBUTIONS

- **Honorable Mention – “Federico Sangati” Best System Across Tasks Award (EVALITA 2026):** Awarded for the paper “*UniTor at DeSegMa-It*”.
- **Program Committee Member:** Reviewer for top ranked NLP conferences like EACL and EMNLP and the Italian Journal of Computational Linguistics.
- **Teaching Assistant:** “Deep Learning”, “Web Mining & Retrieval” and “Fundamentals of Computing” courses (M.S. level), mentoring 50+ students.