LEONARDO ZILLI

leonardo.zilli@studio.unibo.it | leonardozilli.it | github.com/leonardozilli

EDUCATION

2022 - 2025 | International Master's degree in DIGITAL HUMANITIES AND DIGITAL KNOWLEDGE
Alma Mater Studiorum – University of Bologna
GPA: 29.5/30

2019 - 2022 | Bachelor's degree in LANGUAGES IN THE INFORMATION SOCIETY
University of Rome "Tor Vergata"
Final score: 110/110 cum Laude
Thesis: *"Natural Language Processing: History, applications and techniques."*

2021 - 2022 | Bachelor's degree in ENGLISH LANGUAGE, LITERATURE AND CIVILIZATION
Catholic University of Lille, France
Double degree program

WORK EXPERIENCE

09/2024 - current | Research Intern
Alma AI Research Institute, Bologna, Italy
Developing Agentic AI and RAG systems for the legal domain as part of my master thesis' project.

03/2024 - 08/2024 | Student Intern
Alma AI Research Institute, Bologna, Italy
Created and maintained legislative documents datasets through web scraping.
Developed methodologies for the classification and clustering of multilingual documents within the GENAI4LEX project.
The project, conducted with the support of the European Commission funds within ERC HyperModeLex, won the Expression of Interest promoted by the Chamber of Deputies.

ADDITIONAL TRAINING

2024 | LEX2024: LEGALXML STANDARDS, LEGAL ANALYTICS AND AI
School of Law, University of Bologna, Ravenna, Italy

2024 | DIGITAL HUMANITIES & RESEARCH SOFTWARE ENGINEERING SUMMER SCHOOL
University of Edinburgh, Scotland

2023 | WORKSHOP ON OPEN CITATIONS AND OPEN SCHOLARLY METADATA 2023
Department of History, Cultures and Civilizations, University of Bologna, Italy

2023 | 2ND ACM DIGITAL HUMANISM SUMMER SCHOOL
Center for Artificial Intelligence and Machine Learning, TU Wien, Austria

PUBLICATIONS, PROCEEDINGS & CONFERENCES APPEARANCES

Michele Corazza, LEONARDO ZILLI, Monica Palmirani. *Topic Similarity of Heterogeneous Legal Sources Supporting the Legislative Process*, Accepted at CLiC-it 2024.

Presentation. *Metascript*, **TEI 2024**, 7-11 Oct, Buenos Aires, Argentina

Poster. *Metascript*, **AIUCD 2024**, 28-30 May, University of Catania

Presentation. *Metascript*, **Genesis Bologna 2024**, 9-11 May, University of Bologna

Selected Projects

CaseConnect | *Python, transformers, Faiss, Flask, Svelte*

- 2nd place at the Digital Future of Justice hackathon organized by the Council of Europe.
- Led a multidisciplinary team of 4 in designing a system for the semantic comparison of legal cases across different jurisdictions, languages and time periods.
- Implemented legally sound dataset segmentation methodologies, the back-end API for multilingual similarity retrieval and the front-end interface.

FrameTrigger | *Python, PyTorch, transformers, FrameNet*

- Fine-tuned BERT models for Frame Semantic Parsing.
- Achieved state-of-the-art results.

IRIS Dataset Analysis | *Python, Polars, Dask*

- Performed gathering, cleaning and analysis of large scholarly data datasets (up to ~200gb).
- Developed methodologies to map the datasets' coverage.
- The project has been designed in full compliance with Open Science standards and practices.

Skills

Pr. Languages:	Python, JavaScript, Bash
Frameworks and APIs:	Pandas, Dask, transformers, Langchain, Flask
Other Tools:	XML, RDF, Git, Linux, DevOps practices
I speak fluently:	Italian, English [Cambridge C1 certificate], French