Ilaria Salogni

ilariasalogni.github.io @ ilaria.salogni@gmail.com
github.com/IlariaSalogni linkedin.com

I am a final-year MSc student in Natural Language Processing. My research goal is discovering how (and to what extent) linguistic information is used by LLMs to carry out a specific task, leveraging linguistic analysis and statistical modeling. I maintain a list of my publications and research projects on my page.

Major Interests: Explainability; Bertology; Low-resources languages
Other Interests: Machine Translation; Multilinguality; Linguistic continuum

EDUCATION

Present – Sept 2021 | University of Pisa, PISA, Italy
MSc in Human Language Technologies
Major: NLP, Computational Linguistics
Thesis (April 2024): "Hierarchical Quantification: a novel approach to cause of death prevalence estimation in Verbal Autopsies datasets" | Advisor: Fabrizio Sebastiani

Apr 2021 – Sept 2017 | University of Pavia, PAVIA, Italy
Bachelor of Arts ("*Lettere moderne*")
Major: Linguistics
Thesis: "Sociolinguistics of the Italian colonial context: interlingua and lexical borrowings in Amharic" | Advisor: Ilaria Fiorentini

PUBLICATIONS

2023 Salogni Ilaria "Salogni at GeoLingIt: Geolocalization by Fine-tuning BERT" *in Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023) Parma, Italy, September 7th-8th, 2023* [PDF]

EXPERIENCE

Research internship Sept 2023 to Feb 2024
National Research Council (ISTI-CNR) of Pisa, PISA, Italy

- Data sourcing and visualization of text biomedical datasets
- Definition of a NLP system to leverage hierarchical labels to classify textual inputs in prior shift setting

PROJECTS

Through hands-on project I developed strong analytical skills for technical and linguistic issues and to evaluate a range of deep learning models. Among the most relevant ones:

- Geolocating (regression task) sequences exhibiting non-standard language varieties relying solely on linguistic content by fine-tuning two BERT-based and a RoBERTa-based model (Github)
- A quantitative analysis in R of the morphological complexity using French verbal system data (Github)
- Revision of the semi-automatic annotation of the UD-Pipe model on an Italian corpus, to assess domain-adaptation issues

PROGRAMMING AND DATA ANALYSIS SKILLS

- Programming Languages: Python, R
- SW/ Tools: Git, GitHub, LaTeX, Google Colab, Jupyter
- Relevant Libraries: TensorFlow, PyTorch, Keras, NumPy, Pandas, Scikit Learn, ggplot

LANGUAGE SKILLS:

Native Italian speaker
Fluent written and spoken English (IELTS band 8, Feb. 2024).
Fluent written and spoken Spanish