Pietro Lesci

Github: github.com/pietrolesci
LinkedIn: linkedin.com/pietrolesci
Twitter-X: @pietro_lesci
pietrolesci@outlook.com
Personal Page: pietrolesci.github.io

I am a PhD student in Computer Science at the University of Cambridge with a background in econometrics. Broadly, I am interested in Natural Language Processing and Machine Learning. I am passionate about the science of language models: developing methods—also drawing from econometrics—to study the effect of training data on models' behaviour. Currently, I focus on active learning, data valuation, and memorisation estimation.

PROFESSIONAL EXPERIENCE

PhD Candidate & Research Assistant @ University of Cambridge
October 2021 – present | Cambridge, UK

I am working with Prof Andreas Vlachos on studying the effect of training data on the behaviour of language models.

Applied Scientist Intern @ Amazon AWS AI Labs
September 2022 – January 2023 | Barcelona, ES

I worked with Lluis Marquez on efficient dialogue state tracking with large language models.

Senior Associate, Data Science @ Bain & Company
March 2020 – October 2021 | Milan, IT

- Data-driven quality assurance (mining)
- Pricing optimization (packaging machines)
- Predictive maintenance in hydropower plants (energy)
- Client segmentation (road tolls and telco)
- Brand penetration maps (food & beverages)
- Developed an internal Python library for NLP

Research Assistant @ Bocconi University BIDSA Center
August 2019 – March 2020 | Milan, IT

I worked with Prof Dirk Hovy on the identification of online abuse on Twitter and developed the *Wordify* and *MACE* web-apps.

Data Science Trainee @ European Central Bank
February 2018 – November 2018 | Frankfurt am Main, DE

- Data quality management of granular master data
- Database reconciliation as project leader
- Data ingestion and de-duplication

SELECTED PUBLICATIONS

[1] P. Lesci, Y. Fujinuma, M. Hardalov, C. Shang, Y. Benajiba, and L. Marquez. Diable: Efficient dialogue state tracking as operations on tables. ACL 2023 (Findings), 2023.

[2] P. Lesci, C. Meister, T. Hofmann, A. Vlachos, and T. Pimentel. Causal estimation of memorisation profiles. ACL 2024.

[3] P. Lesci and A. Vlachos. Computationally efficient active learning for large and imbalanced datasets. NAACL 2024.

AWARDS & INVITED TALKS

Awarded the Translated's *"Imminent Research Grant"*.
Invited talk @ Ammagamma, 2022: *"Transformer Models in NLP"*.

EDUCATION

MSc Economic and Social Sciences @ Bocconi University
Sept 2016 - Mar 2019 | Milan, IT

Bayesian Statistics and Econometrics. Dissertation: *"Deep Learning: A Statistical Perspective"*. Grade: 104/110

BSc Economics and Management @ Università Cattolica del Sacro Cuore
Sept 2013 - Sept 2016 | Milan, IT

Macroeconomics and Agent-Based Models. Dissertation: *"Market Sentiment and Monetary Policy"*. Grade: 110/110 cum laude

Scientific High School Diploma
Sept 2008 - July 2013 | Castrovillari, IT

Liceo Scientifico. Grade: 95/100

ACADEMIC ROLES

Supervisor for MSc Thesis Projects
Co-supervised two MSc students with Prof Andreas Vlachos and Prof Pietro Lio.

Head Demonstrator & Supervisor
Machine Learning and Real-world Data, University of Cambridge, 2022 and 2023.

Conference Reviewer
Reviewed for top-ranked international conferences (ACL, EACL, and EMNLP).

SKILLS

Natural & Formal Languages
Italian (native), English (fluent), and French (basic). Python, R, Stata, SQL, LaTeX, Julia, MATLAB, and Rust.

EXTRACURRICULAR

Music Education & Practice
Enrolled in the BSc in Jazz, GPA 29.7/30. Classical Piano (12 years) and Bass Guitar (10 years) studies obtaining Sol-Fa degree and Piano 5th year diploma.