

LUCA MALAGUTTI

Hofwiesenstrasse 376, 8050, Zurich

🏠 lucamalagutti.com

✉ malagutti.luca317@gmail.com

🌐 linkedin.com/in/LucaMalagutti

🔊 LucaMalagutti

Work Experience

Thomson Reuters Labs

Applied Scientist

Apr. 2024 – Present

Zug, Switzerland

- Applied Research Intern (Apr. 2024 – Oct. 2024)

ETH Zurich

Research Assistant

Oct. 2023 – Mar. 2024

Zurich, Switzerland

- Research assistant under the supervision of Prof. Ryan Cotterell.

SEI Novus

Data Science Intern

Mar. 2022 – Sep. 2022

Zurich, Switzerland

- Built and deployed from scratch a complete NLP pipeline for automatic data extraction from financial documents.

Education

ETH Zurich

MSc in Computer Science

Sep. 2020 – Aug. 2023

Zurich, Switzerland

- Overall grade: 5.45/6.0
- Selected courses: Natural Language Processing, Advanced Machine Learning, Machine Perception, Big Data.

Politecnico di Milano

BSc in Computer Engineering

Sep. 2017 – Jul. 2020

Milan, Italy

- Final grade: 110/110
- Selected courses: Probability and Statistics, Linear Algebra, Algorithms and Principles of Computer Science.

Teaching Experience

Natural Language Processing, Large Language Models

Teaching Assistant

2021, 2022, and 2023

ETH Zurich

- Instructors: Prof. Ryan Cotterell, Prof. Mrinmaya Sachan, Prof. Florian Tramèr, and Prof. Ce Zhang
- Wrote lecture notes, held lectures and tutorials, prepared and marked assignments and exam questions over four course editions.

Publications

- Mario Giulianelli, **Luca Malagutti**, Juan Luis Gastaldi, Brian DuSell, Tim Vieira, and Ryan Cotterell. "On the Proper Treatment of Tokenization in Psycholinguistics." *EMNLP 2024* (Long Paper)
- Juan Luis Gastaldi, John Terilla, **Luca Malagutti**, Brian DuSell, Tim Vieira, and Ryan Cotterell. "The Foundations of Tokenization: Statistical and Computational Concerns." *Under review* (Preprint)
- **Luca Malagutti**, Andrius Buinovskij, Anej Svete, Clara Meister, Afra Amini, and Ryan Cotterell. "The Role of n -gram Smoothing in the Age of Neural Networks." *NAACL 2024* (Long Paper)
- Clara Meister, Tiago Pimentel, **Luca Malagutti**, Ethan Wilcox, and Ryan Cotterell. "On the Efficacy of Sampling Adapters." *ACL 2023* (Long Paper)
- Afra Amini, Ryan Cotterell, John Hewitt, **Luca Malagutti**, Clara Meister, and Tiago Pimentel. "Generating Text from Language Models." *ACL 2023* (Tutorial)

Awards

Best Freshmen Award

2018

Politecnico di Milano

- Awarded monetary prize granted to the best-performing freshmen of the 2017-2018 academic year.

University Tuition Scholarship

2018, 2019, and 2020

Politecnico di Milano

- Awarded total (2018) and partial (2019, 2020) tuition reduction granted to high-performing students.

Skills

Programming Languages	Python, R, Java, C, SQL, L ^A T _E X
Tools and Frameworks	NumPy, Pandas, SciKit, PyTorch, Huggingface, AWS, Git, Docker
Languages	Italian (native), English (C2 level), German (B1 level)

Selected Projects

Divergences for Natural Language Generation

Oct. 2022 – Apr. 2023

Master's Thesis

Rycolab, ETH Zurich

- Supervision by Ph.D. student Clara Meister and Prof. Ryan Cotterell.
- Explored the use of divergence functions as alternative fine-tuning losses for Natural Language Generation models.

Word Embedding initialization for Knowledge Graph Link Prediction

Feb. 2022 – Jun. 2022

Semester Project

ETH AI Center

- Supervision by Dr. Carl Allen.
- Experimented with improving performance and convergence speed of knowledge graph models through the use of pre-trained neural word embeddings.

Automatic Textbook Summarization

Sep. 2021 – Jan. 2022

Semester Project

NLPED Lab, ETH Zurich

- Supervision by Ph.D. student Kumar Shridhar and Prof. Mrinmaya Sachan.
- Created, analyzed, and modeled a novel dataset for downstream tasks in educational NLP.

Deep-Mon

Feb. 2020 – Jun. 2022

Semester Project

NECST Lab, Politecnico di Milano

- Supervision by Dr. Rolando Brondolin and Prof. Marco Santambrogio.
- Improved an eBPF-based container monitoring tool by implementing its memory, disk, and I/O monitoring capabilities, and creating a new CLI.