

LUCA MALAGUTTI

🏠 lucamalagutti.com

✉ malagutti.luca317@gmail.com

🌐 linkedin.com/in/LucaMalagutti

🔗 [LucaMalagutti](https://github.com/LucaMalagutti)

Work Experience

Thomson Reuters Labs

Apr. 2024 – Present

Applied Scientist

Zug, Switzerland

- Leading AI adoption at Reuters News, coordinating multiple workstreams to integrate Generative AI in editorial workflows.
- Designed and developed agentic systems, specialized language models, and AI-powered financial reporting tools, improving speed and accuracy of editorial output.

ETH Zurich

Oct. 2023 – Mar. 2024

Research Assistant

Zurich, Switzerland

- Research Assistant under the supervision of Prof. Ryan Cotterell. Conducted research in tokenization theory, sampling methods, and language model generation, co-authoring papers at ICLR, ACL, EMNLP, and NAACL (50+ citations).

SEI Novus

Mar. 2022 – Sep. 2022

Data Science Intern

Zurich, Switzerland

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

Education

ETH Zurich

Sep. 2020 – Aug. 2023

MSc in Computer Science

Zurich, Switzerland

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

Politecnico di Milano

Sep. 2017 – Jul. 2020

BSc in Computer Engineering

Milan, Italy

- Final grade: 110/110
- Selected courses: Probability and Statistics, Linear Algebra, Algorithms and Principles of Computer Science.
- Awarded Best Freshmen Award and merit-based tuition Scholarships.

Teaching Experience

Natural Language Processing, Large Language Models

2021, 2022, and 2023

Teaching Assistant

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

- Juan Luis Gastaldi, John Terilla, **Luca Malagutti**, Brian DuSell, Tim Vieira, and Ryan Cotterell. "The Foundations of Tokenization: Statistical and Computational Concerns." *ICLR 2025* (Long Paper)
- 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)
- **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)

Skills

Programming Languages
Machine Learning & NLP
Gen AI & LLMs
MLOps
Languages

Python, R, Java, C, SQL, \LaTeX
PyTorch, HuggingFace Transformers, spaCy, scikit-learn, pandas
Fine-tuning, prompt engineering, evals, SLMs, agents
AWS, Docker, Git, CI/CD, experiment tracking, data pipeline design
Italian (native), English (C2 level), German (B1 level)