

Gabriele Sarti

PHD STUDENT IN NATURAL LANGUAGE PROCESSING, UNIVERSITY OF GRONINGEN

✉ g.sarti@rug.nl | 🏠 gsarti.com | 📧 gsarti | 📺 gabrielesarti | 🐦 @gsarti_ | 🗣️ Gabriele Sarti

Education

University of Groningen, Center for Language and Cognition

Groningen, Netherlands

PhD in Natural Language Processing

September 2021 - Present

- Working on interactive interpretable neural machine translation, supervised by Arianna Bisazza, Malvina Nissim and Grzegorz Chrupala.
- Part of the NWO-funded network “InDeep: Interpreting Deep Learning Models for Text and Sound”.

University of Trieste, International School for Advanced Studies (SISSA)

Trieste, Italy

Master of Science (MSc) in Data Science and Scientific Computing

October 2018 - December 2020

- Data Science for Social Sciences track, all courses were taught and assessed in English. Graduation score: 110 Cum Laude.
- Thesis: “Interpreting Neural Language Models for Linguistic Complexity Assessment”. Supervisors: Felice Dell’Orletta, Davide Crepaldi.

Cégep of Saint-Hyacinthe

Saint-Hyacinthe, Canada

Collegial Studies Degree (DEC) in Management Informatics

August 2015 - June 2018

- Graduated with work-study mention with final internship at Skytech Communications under the supervision of Prof. Hugo St-Louis.
- All courses are taught and assessed in French. Final R Score: 34.13.

Work Experience

Aindo

Trieste, Italy

Research Scientist

November 2020 - August 2021

- Sequence-to-sequence question answering applied to structured predictions for the clinical domain.
- Experimental work in cross-linguistic transfer to Italian for neural language models pre-trained in English.
- Temporally-aware multi-interaction neural recommender systems for online clothes resellers.
- Controllable sketch-to-image synthesis for the fashion industry with StyleGANs and CLIP.

Institute of Computational Linguistics “Antonio Zampolli” (ILC-CNR)

Pisa, Italy

Research Assistant Intern in Natural Language Processing

October 2019 - December 2019

- Hosted by Felice Dell’Orletta and Dominique Brunato at the ItaliaNLP Lab. Development of a multitask training pipeline for neural language models leveraging text and human cognitive processing signals. Models were evaluated on complexity assessment tasks and compared by means of interpretability approaches like probing tasks, RSA and PWCCA.

Skytech Communications Inc.

Montréal, Canada

Machine Learning Engineer Intern

February 2018 - June 2018

- Implemented and deployed a solution based on convolutional neural networks to augment the performances of an automatic record keeping systems by adding handwritten characters recognition capabilities.
- Development of a software platform using sentiment analysis and information retrieval techniques in order to highlight inappropriate student comments in professors’ evaluation forms.

Talks and Publications

SELECTED PUBLICATIONS

🔗 **Contrastive Language-Image Pre-training for the Italian Language**

Federico Bianchi, Giuseppe Attanasio, Raphael Pisoni, Silvia Terragni, Gabriele Sarti, Sri Lakshmi

Arxiv preprint, 2021

🔗 **That Looks Hard: Characterizing Linguistic Complexity in Humans and Language Models**

Gabriele Sarti, Dominique Brunato, Felice Dell’Orletta

Hon. mention, Best student paper, Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2021)

ETC-NLG: End-to-end Topic-Conditioned Natural Language Generation

Ginevra Carbone, Gabriele Sarti

Italian Journal of Computational Linguistics (IJCoL), Issue 6-2, 2020

Italian Transformers Under the Linguistic Lens

Alessio Miaschi, Gabriele Sarti, Dominique Brunato, Felice Dell'Orletta, Giulia Venturi

Hon. mention, Best student paper, *Proceedings of the 7th Italian Conference on Computational Linguistics (CLIC-it 2020)*

Improving Complexity and Acceptability Prediction with Multi-task Learning on Self-Supervised Annotations

Gabriele Sarti

Hon. mention, Best student paper, *Proceedings of the 7th Evaluation Campaign of NLP and Speech Tools for Italian (EVALITA 2020)*

SELECTED TALKS

Empowering Human Translators with Interpretable Interactive Neural Machine Translation

Gabriele Sarti — *Oral presentation, A glimpse of the future track, XAI4Debugging Workshop at NeurIPS 2021*

Characterizing Linguistic Complexity in Humans and Language Models

Gabriele Sarti — *Invited talk at the MilaNLP group at Bocconi University, IT*

Neural Language Models: The New Frontier of Natural Language Understanding

Gabriele Sarti — *Young Chapter of Italian Statistical Society (ySIS), StaTalk 2019*

The Educational Impact of Artificial Intelligence

Gabriele Sarti, Alexandre Brunet — *38th Symposium of Québec Association for Collegial Pedagogy (AQPC), 2018*

Extracurricular Activity

SCIENTIFIC COMMUNICATION

Italian Association for Computational Linguistics (AILC)

Communication Team Member

Italy

January 2020 - Present

- Promote the field of computational linguistics at local and national scientific festivals through presentations and interactive demonstrations.

Trieste Next Scientific Festival

System Demonstration Organizer

Trieste, Italy

June 2019 - September 2019

- Leveraged the first publicly available GPT-2 implementation to build a fake historical letter generator, and organized the 3-day live demonstration "Altalo Svevo: Letters from an Artificial Intelligence" with more than 3000 participants aimed at sharing information about uses and limitations of natural language generation systems.

RESEARCH PROJECTS

March 2020  [COVID-19 Semantic Browser](#), in collaboration with Area Science Park and AILC

SOCIAL ENGAGEMENTS

2020 President and Founder,  [Artificial Intelligence Student Society \(AI2S\)](#)

2015-2020 Volunteer, AFS Italy and AFS Canada

Honors & Awards

2018 Excellence Scholarship for Data Science Students, SISSA

Trieste, Italy

2018 Valedictorian of the Management Informatics Class of 2018, Cégep of Saint-Hyacinthe

Saint-Hyacinthe

2016 Perseverance Scholarship for Informatics Students, Hydro-Québec

Saint-Hyacinthe

2013 Full Merit Scholarship towards an AFS Exchange Year in Canada, TIM

Rome, Italy

Academic Referees

Malvina Nissim

Full Professor at the Center for Language and Cognition of the University of Groningen — m.nissim@rug.nl

Luca Bortolussi

Full Professor and Program Coordinator of the Data Science MSc, University of Trieste — lbortolussi@units.it