

Silvia Casola

Contact:

- s.casola@lmu.de
- +39 3277408493

Personal details:

- Place of birth: Mazara del Vallo (TP), Italy
- Date of birth: 08 July 1993
- Citizenship: Italian

Updated: June 9, 2025

Research Interests

Natural Language Generation, Summarization, Simplification, Human Label Variation, Evaluation, Natural Language Processing, Deep Learning, Machine Learning.

Research Profile

I am a post-doc researcher at the Ludwig Maximilian University of Munich, under the supervision of Prof. Barbara Plank. Previously, I was a post-doc researcher at the University of Turin (PI: Dr. Valerio Basile), where I worked on perspective-aware NLP, disagreement and human label variation (Amazon-Science granted project).

I was awarded a Ph.D. cum laude from the University of Padua with a grant from Fondazione Bruno Kessler (Advisor: Alberto Lavelli); my main focus was on summarization and simplification, mostly in the patent domain. During my Ph.D., I visited Universitat Pompeu Fabra under the supervision of Prof. Horacio Saggion and interned remotely with Spotify Research and Huawei Research.

Previously, I obtained a Laurea Magistrale (Master's Degree) with honors in Computer Engineering & Data Science from Politecnico di Torino and then worked as a research assistant at the same University.

My research interests lie in NLP, focusing on Natural Language Generation.

Academic Appointments

- **Ludwig Maximilian University of Munich**, Munich, Germany *September 2024 – present*
and **Munich Center for Machine Learning (MCML)**
Post-doctoral Researcher
Advisor: Prof. Barbara Plank
 - **University of Turin**, Turin, Italy *May 2023 – August 2024*
Post-doctoral Researcher
Advisor: Dr. Valerio Basile
Project granted by Amazon Science
I worked on perspectivism in Natural Language Processing. Our approaches exploit and analyze the disagreement among annotators rather than resolve it, e.g., through majority voting.
 - **Fondazione Bruno Kessler**, Trento, Italy *February – May 2023*
Research Fellow
Advisor: Alberto Lavelli
I worked on summarization and simplification with a focus on patents.
 - **Politecnico di Torino**, Turin, Italy *January – October 2019*
Research Assistant
Advisor: Prof. Silvia Chiusano
I was part of a team designing a recommendation system for health structures (CANP project).
-

Non-academic Work Experience

- **aequa-tech**, Remote *March – October 2024*
NLP Research Scientist
As part of the "Friends of aequa-tech" Fellowship program
My work focused on using summarization in the news domain.
- **Spotify Research**, Remote *June – September 2023*
NLP Research Scientist Intern
My work focused on using LLMs to generate summaries of books.

- **Teleperformance Rome**, Remote June – July 2023
Python instructor and trainer
 - **Huawei Research Ireland**, Remote July – December 2021
NLP Ph.D. Intern
I developed a module to correct misspelled queries in the AppGallery, which has been deployed. Research-wise, I worked on ensuring factuality and minimizing hallucination in a sequence-to-sequence system for news summarization.
-

Education

- **University of Padua**, Padua, Italy October 2019 – September 2023
Ph.D. in Brain, Mind and Computer Science
cum laude and the additional label of Doctor Europeus
with a grant from Fondazione Bruno Kessler (~ 45K€)
Advisor: Alberto Lavelli (FBK)
Thesis: *Natural Language Processing for technology foresight – Summarization and simplification: the case of patents*
 - **Politecnico di Torino**, Turin, Italy 2016 – 2018
Laurea Magistrale (2 years M.S. equivalent) in Computer Engineering (Data Science curriculum)
Thesis: *Reducing waiting times and crowding in hospital emergency departments using Machine Learning*
Supervisors: Prof. Ricard Gavaldà (UPC) and Prof. Silvia Chiusano (Politecnico di Torino)
Grade: 110/110 cum laude
 - **Politecnico di Torino**, Turin, Italy 2013 – 2016
Laurea (3 years B.S. equivalent) in Computer Engineering.
Grade: 102/110
 - **Robert Gordon University**, Aberdeen, UK 2012 – 2013
Certificate of Higher Education in Journalism
-

Schools and short courses

- **University of Bari**, Bari, Italy June 2024
AILC Lectures on Computational Linguistics 2024
 - **Technische Universität München (TUM)**, Munich, Germany August 2021
1st Munich Legal Tech Summer School
 - **University of Grenoble and Naver Labs Europe**, Online January 2021
Advanced Language Processing Winter School (ALPS)
 - **coursera.org**
Patent Law; by Prof. Polk Wagner (University of Pennsylvania)
Deep Learning Specialization; by Prof. Andrew Ng (Stanford University) - *Certificate*
Machine Learning Course; by Prof. Andrew Ng (Stanford University)
-

Visiting Periods

- **Universitat Pompeu Fabra**, Barcelona, Spain March – August 2022
Visiting Ph.D. Student
I visited the TALN group under the supervision of Prof. Saggion. I collaborated with the group researchers and developed my interest in summarization and simplification.
 - **Universitat Politècnica de Catalunya**, Barcelona, Spain September 2017 – July 2018
Erasmus Program Master's student – MS in Artificial Intelligence, MS in Innovation and Research.
I followed courses in the Machine Learning area and wrote and defended my Master's thesis.
-

Awards

- **Marie Skłodowska-Curie Actions Seal of Excellence** 2025
Proposal: PERSEO - PERsonalized text Simplification and ElabOration (HORIZON-MSCA-2024-PF-01)
 - **Outstanding Paper Award** ACL 2024
MultiPiCo: Multilingual Perspectivist Irony Corpus, Annual Meeting of the Association for Computational Linguistics
-

Media coverage

- Our project on understanding the use of GenAI tools in Italy was covered in the national newspaper "Corriere della Sera" in the article "Cosa fanno gli italiani con ChatGpt, Gemini & co? Un sondaggio per indagare l'uso (reale) dell'AI. E aiutare la ricerca" ([link](#)).
 - Our paper *MultiPiCo: Multilingual Perspectivist Irony Corpus* was covered in the magazine "Otto: discorsi diretti" of the University of Torino in an interview to Dr. Valerio Basile ([link](#)).
-

Teaching

- **NLG Evaluation seminar** April-July 2025
Course formally assigned to me and Dr. Liu as part of the Master's in Computational Linguistics 2h/week for 14 weeks. I covered classes about automatic evaluation (standard metrics, LLMs as evaluators, data contamination, etc)
 - **Summarization seminar** April-July 2025
Course formally assigned to me as part of the Master's in Computational Linguistics . 2h/week for 14 weeks. The course was completely designed, prepared, and taught by me. The project covered open topics in summarization and GenAI (hallucinations, agents, personalization, multimodality, etc).
 - **HumanAIze – Master II Livello** February 2025
Lectures and laboratory sessions on NLP within the context of the HumaAIze Master's program (Politecnico di Torino and University of Turin), aimed at providing humanities graduates with AI and other digital competencies. 10 hours.
 - **Summarization and Simplification Seminar** October 2024 – January 2025
Course formally assigned to me as part of the Master's in Computational Linguistics. 2h/week for 14 weeks. The course was completely designed, prepared, and taught by me. The course covered topics in summarization and simplification, and the evaluation was project-based.
 - **Vertiefung der Grundlagen der Computerlinguistik** October 2024 – present
Course formally assigned to me and Prof. Plank as part of the Master's in Computational Linguistics and co-taught by me. The course consisted of lectures (2h/week) and exercises (2h/week) for 14 weeks. Lectures cover standard NLP foundation topics (e.g., BPE tokenization, annotation, neural networks, word vectors, CNN, RNN, attention, transformers). The course is co-taught by me, Prof. Plank and MA Säuberli.
 - **Trustworthy Data-centric AI** October 2024 – present
Course formally assigned to me and Dr. Liu and co-taught by me. The course consisted of lectures (2h/week) and exercises (1h/week) for 14 weeks. The course covers advanced concepts in NLP, e.g., human annotation and crowdsourcing, silver data, FAIR principles, model behavioral testing, mechanism interpretability, human and automatic evaluation, LLMs as judges. The course is co-taught by me, Prof. Plank, Dr. Liu, Dr. Peng., Dr. Hedderich and MA Säuberli.
 - **Programming Teaching Assistant (Python)** October 2020 – February 2021
Lecturer, exerciser, and homework grader; first-year Math undergrads at the University of Padua.
 - **Computer Science Course Teaching Assistant (C)** September 2017 – February 2018
Exerciser and lab assistant; first-year Engineering undergrads at Politecnico di Torino.
-

Supervision

- **Official Master's thesis supervision**
 - Zhaochen Guo, 2024 (LMU, advisor)
 - Nils Grünefeld, 2019 (LMU, advisor)
 - **Official Bachelor thesis supervision**
 - Yueou Li, 2024 (LMU, advisor)
-

Academic service

- **Shared task organizer**
LeWiDi - Learning With Disagreement (EMNLP 2025)
- **Mentor**
ACL Student Research Workshop (2024)
- **Ph.D. course student representative** , Padua, Italy
University of Padua
February 2022 – February 2023
- **Reviewer / PC member**
 - Nature Scientific Reports
 - IEEE Transactions on Affective Computing
 - Neurocomputing
 - Natural Language Engineering
 - ACL Rolling Review (ARR)
 - ACM/SIGAPP Symposium On Applied Computing, 2025
 - Generation, Evaluation & Metrics Workshop (GEM), 2022, 2023, 2025
 - Mind the AI Gap Workshop, 2025
 - European Conference on Artificial Intelligence (ECAI), 2024
 - ACL Student Research Workshop (ACL SRW), 2023, 2024
 - Workshop on Natural Language for Artificial Intelligence @ AixIA, 2024
 - EACL Student Research Workshop (EACL SRW), 2024
 - NLPerspectives Workshop, 2024
 - Italian Conference on Computational Linguistics (CliC-It), 2023, 2024
 - Empirical Methods in Natural Language Processing (EMNLP), 2023
 - European Symposium on Artificial Neural Networks (ESANN), 2020
 - Social Media Mining for Health Applications Workshop (#SMM4H), 2020
 - Qeios

Invited talks and presentations

- *Large Language Models in a polyphonic world* 2024
Munich Center for Machine Learning Industry pitch talks, Munich RE, Munich
- *MultiPICO: Multilingual Perspectivist Irony Corpus* 2024
Annual Meeting of the Association for Computational Linguistics (ACL), Bangkok
- *Confidence-based Ensembling of Perspective-aware Models* 2024
Workshop on Perspectivist Approaches to NLP (NLPerspectives), LREC-COLING, Turin
- *A perspectivist benchmark for training and evaluation* 2024
2nd POPULUS Workshop, University of Turin
- *Confidence-based Ensembling of Perspective-aware Models* 2023
Empirical Methods in Natural Language Processing (EMNLP), Singapore
- *Testing ChatGPT for Stability and Reasoning* 2023
Italian Conference on Computational Linguistics (CLiC-it), Venice
- *Creating a Silver Standard for Patent simplification* 2023
Association for Computing Machinery's Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval (SIGIR), Online
- *Benchmarking Natural Language Processing Algorithms for Patent Summarization* 2023
Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech), SIGIR, Online
- *Computers and Language* 2023
Istituto di Istruzione Superiore Evangelista Torricelli, Milan
- *What's in a (dataset's) name? The case of BigPatent* 2022
Workshop on Generation, Evaluation & Metrics, (EMNLP), Online
- *WITS: Wikipedia for Italian Text Summarization* 2021
Italian Conference on Computational Linguistics (CLiC-IT), Milan

- *Natural Language Processing: an overview* 2021
Machine Learning Course, the University of Padua (invited by Prof. F. Aiolli)
 - *FBK@SMM4H2020: RoBERTa for detecting medications on Twitter* 2020
Social Media Mining for Health Applications Workshop & Shared Task (COLING), Online
-

Participation in Research projects

- *POPULUS: Multilingual Perspective-aware NLU* 2023 - 2024
Role: Post-doc researcher
Funding: Amazon Science
PI: Valerio Basile (University of Turin).
 - *eCREAM: enabling Clinical Research in Emergency and Acute care Medicine through automated data extraction* 2023
Role: External researcher
Funding: European Commission (Horizon Europe)
PI: Alberto Lavelli (Fondazione Bruno Kessler).
 - *CANP: La Casa nel Parco* 2019
Role: Assistant Researcher
Funding: Regione Piemonte
PI: Silvia Chiusano (Politecnico di Torino).
-

Other activities

CD: 50/50, co-founder

Non-profit association for promoting gender equality and diversity in STEM.

The association is part of Repubblica Digitale and sponsored by ASviS - Alleanza Italiana per lo Sviluppo Sostenibile. Some of my activities were:

- Liceo Scientifico Morgagni, Roma: course (PCTO) on Python and data literacy (2022, 40h)
- Istituto di Istruzione Superiore Evangelista Torricelli, Milano: Talk on Computers and language (2023)
- Istituto di Istruzione Superiore Piero Gobetti, Torino: course (PCTO) on AI and NLP (2024, 20hx3 classes)

Direction Under30, coordination, communication and organization

Coordination and organization of a project to select, discuss, and award theatrical projects in the context of the national DU30 Festival (2018 — Present).

Technical Skills

Programming languages: Python (first language), SQL, C. Basics in R, C++, and Java.

Machine/Deep Learning stack: Pytorch; Numpy, Pandas, Matplotlib, Scipy, Scikit-learn, etc.; LangChain & Hugging Face ecosystems, NLTK, Spacy; Weight and Biases. Experience with large-scale GPU clusters (slurm).

Languages

Italian (native)

English (full professional proficiency) - [C2 Certificate \(10/2020\)](#)

Spanish (conversational proficiency)

List of Publications

- **Casola***, Liu*, Peng*, Gatt, Plank (2025): Evaluation Should Not Ignore Variation: On the Impact of Reference Set Choice on Summarization Metrics, submitted to ARR.
- Lo, **Casola***, Sezeren, Basile, Sansonetti, Uva, Bernardi (2025): PERSEVAL: A Framework for Perspectivist Classification Evaluation, submitted to ARR.
- Du, Mordorf, **Casola***, Yao, Litschko, Plank (2025): Reason to Rote: Rethinking Memorization in Reasoning, submitted to ARR.

- Akhtar*, **Casola***, Madeddu*, Basile, Patti (2024): Whose Opinion Matters? Target-Centered Perspectivist Modelling of Abusive Language, submitted to Affective Computing.
- Balestrucci*, **Casola***, Lo*, Basile, Mazzei (2024): I'm sure you're a real scholar yourself: Exploring Ironic Content Generation by Large Language Models, Empirical Methods in Natural Language Processing Findings (EMNLP Findings) 2024.
- **Casola***, Frenda*, Lo*, Sezerer, Uva, Basile, Bosco, Pedrani, Rubagotti, Patti, Bernardi (2024): MultiPICO: Multilingual Perspectivist Irony Corpus, Annual Meeting of the Association for Computational Linguistics (ACL) 2024 - **ACL 2024 Outstanding Paper**
- Bourgeade, **Casola**, Wizan, Bosco (2024): Data Augmentation through Back-Translation for Stereotypes and Irony Detection, Italian Conference on Computational Linguistics (CLiC-it) 2024.
- Basile, **Casola**, Frenda, Lo (2024): PERSEID-Perspectivist Irony Detection: A CALAMITA Challenge, Italian Conference on Computational Linguistics (CLiC-it) 2024.
- Frenda, Piergentili, Savoldi, Madeddu, Rosola, **Casola**, Ferrando, Patti, Negri, Bentivogli (2024): GFG-Gender-Fair Generation: A CALAMITA Challenge, Italian Conference on Computational Linguistics (CLiC-it) 2024.
- **Casola**, Lo, Basile, Frenda, Cignarella, Patti, Bosco (2023): Confidence-based Ensembling of Perspective-aware Models. Empirical Methods in Natural Language Processing (EMNLP) 2023
- **Casola**, Lavelli, Saggion (2023): Creating a Silver Standard for Patent simplification. Association for Computing Machinery's Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval (SIGIR) 2023
- **Casola**, Labruna, Lavelli, Magnini (2023): Testing ChatGPT for Stability and Reasoning: A Case Study Using Italian Medical Specialization Tests. Italian Conference on Computational Linguistics (CLiC-it) 2023
- Frenda, Lo, **Casola**, Scarlini, Marco, Basile, Bernardi (2023): Does anyone see the irony here? Analysis of perspective-aware model predictions in irony detection. Workshop on Perspectivist Approaches to NLP (NLPerspective), ECAI 2023
- **Casola**, Lavelli (2023): Benchmarking Natural Language Processing Algorithms for Patent Summarization. 3rd Workshop on Patent Text Mining and Semantic Technologies, SIGIR 2023
- **Casola**, Lavelli, Saggion (2022): What's in a (dataset's) name? The case of BigPatent. Workshop on Generation, Evaluation & Metrics (GEM), EMNLP 2022
- Obonyo, **Casola**, Saggion (2022): Exploring the limits of a base BART for multi-document summarization in the medical domain. Workshop on Scholarly Document Processing, COLING 2022
- **Casola**, Lavelli (2022): Summarization, Simplification, and Generation: The Case of Patents. Expert Systems with Applications
- **Casola**, Lauriola, Lavelli (2022): Pre-Trained Transformers: An Empirical Comparison. Machine Learning with Applications
- **Casola**, Lavelli (2021): WITS: Wikipedia for Italian Text Summarization. Italian Conference on Computational Linguistics (CLiC-IT) 2021
- Louvan, **Casola**, Magnini (2021): Investigating Continued pretraining for Zero-Shot Cross-Lingual Spoken Language Understanding. Italian Conference on Computational Linguistics (CLiC-IT) 2021
- **Casola**, Lavelli (2020): FBK@SMM4H2020: RoBERTa for detecting medications on Twitter. Social Media Mining for Health Applications Workshop & Shared Task, COLING 2020
- Bassi, **Casola**, Mancinelli, Lai, Salcuni (2020): Internalizing-Externalizing Symptoms as Predictors of Problematic Smartphone Use among Adolescents: A Machine Learning approach. International Congress of Clinical and Health Psychology in Children and Adolescents.
- Bazzano, Montuschi, Lamberti, Paravati, **Casola**, Ceròn, Londoño, Tanese (2017): Mental Workload Assessment for UAV Traffic Control Using Consumer-Grade BCI Equipment. Intelligent Human Computer Interaction (IHCI) Conference 2017.