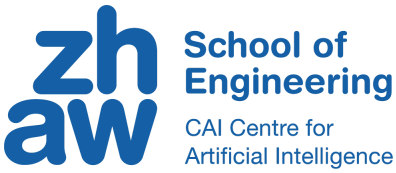
SwissText 2025

Proceedings of the 10th Edition of the Swiss Text Analytics Conference

13–15 May, 2025
Swiss Association for Natural Language Processing
Winterthur, Switzerland

Presenting Organizations



Co-Organizers of Swiss NLP Days

Gold Sponsor

Bronze Sponsors

Partners

data innovation alliance

Fachhochschule Graubünden

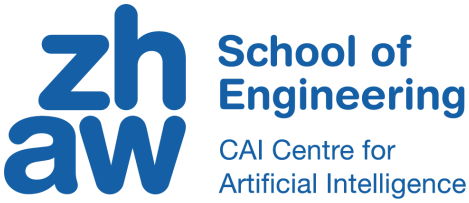
Fernfachhochschule
Schweiz
Mitglied der SUPSI

HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

unine
Université de Neuchâtel

HSLU Hochschule Luzern

Universität Zürich UZH


zh aw
School of Engineering
CAI Centre for Artificial Intelligence

datalab
www.zhaw.ch/datalab

zh aw
School of Applied Linguistics
ILC Institute of Language Competence

Datalab – The ZHAW Data Science Laboratory

SGAICO

SD:>_
[swissdevjobs.ch]

WIN LINK
INFORMATION AND COMMUNICATION TECHNOLOGY

swiss made software

COMMUNITY PARTNER
NLP ZURICH

alp ict
western switzerland digital cluster

©2025 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-950737-19-2

Introduction

In 2025, the SwissText conference celebrated its 10th anniversary! To mark this milestone, we organized a special anniversary edition:

- SwissText 2025 returned to the Zurich University of Applied Sciences (ZHAW) in Winterthur, where SwissText was first held in 2016
- We invited selected Keynote Speakers from previous years to share their (updated!) insights on the current status of NLP
- We extended SwissText with two new events:
  - A "Crash Course" in NLP, held biweekly in 2025, designed to introduce Natural Language Processing to beginners
  - The "NLP Expo 2025", a business fair that welcomed more than 350 participants where companies, startups and research groups showcased their NLP & GenAI solutions

These three events – SwissText, Expo and Crash Course – were combined under the umbrella of the "Swiss NLP Days 2025", jointly organized by SwissNLP and Zurich University of Applied Sciences, with the support of numerous partners, sponsors and co-organizers.

The 10th SwissText conference took place at the Eulachpassage of the Zurich University of Applied Sciences (ZHAW) in Winterthur from 14th to 15th May. Around 100 participants attended keynotes, talks, workshops, and the poster session. Our call for papers asked for contributions in two major tracks: Scientific and Impact. In addition, we introduced a new "Corpora Track", which focused on presenting Swiss datasets. We received a total of 35 submissions, of which we selected 8 for oral presentation, and 23 as posters.

In addition, there were four workshops on the following topics:

- LiRI Corpus Platform
- Battle of NLP Ideas
- Instruction fine-tuning & QLoRA
- LLM Post-Training & DPO

These proceedings include:

- 13 full papers from the Scientific Track
- 3 papers from the Corpora Track
- 16 abstracts from the Impact Track
- Summaries and contributions from the workshops and shared tasks.

We would like to thank our keynote speakers Claudiu Musat (Google DeepMind), Marco Passarotti (Università Cattolica del Sacro Cuore), Iryna Gurevych (Technical University of Darmstadt), Eneko Agirre (University of the Basque Country), Holger Schwenk (Meta AI), and Margot Mieskes (University of Applied Sciences Darmstadt). Their insights and contributions were much appreciated.

We are also deeply grateful to our sponsors and partners who supported us. In addition, we would like to express our sincere gratitude to everyone who has contributed to this conference for their excellent work.

Your sincerely,

Mark Cieliebak, Don Tuggener, and Manuela Hürlimann

Winterthur

October 2025

General Chairs:

Prof. Dr. Mark Cieliebak - Centre for Artificial Intelligence

Program Chairs:

Dr. Don Tuggener - Centre for Artificial Intelligence
Manuela Hürlimann - Centre for Artificial Intelligence

Organizers:

Jasmin Heierli - Institute of Business Information Technology
Jonathan Gerber - Intelligent Information Systems
Jasmin Saxer - Intelligent Information Systems

Program Committee:

Fernando Benites, Pascual Cantos-Gómez, Ernesto William De Luca, Serge Heiden, Valia Kordoni, Brigitte Krenn, Alexander Mehler, Margot Mieskes, Andrei Popescu-Belis, Georg Rehm, Fabio Rinaldi, Tanja Samardžić, Yves Scherrer, Helmut Schmid, Sabine Schulte Im Walde, Manfred Stede, Ludovic Tanguy, Don Tuggener, Juergen Vogel, Manfred Vogel, Egon Werlen, Torsten Zesch

Invited Speakers:

Dr. Claudiu Musat, Google DeepMind

Prof. Dr. Marco Passarotti, Università Cattolica del Sacro Cuore, Milan, Italy

Prof. Dr. Iryna Gurevych, Technical University of Darmstadt (Germany), co-affiliated with Mohamed bin Zayed University of Artificial Intelligence in Abu Dhabi (UAE) and INSAIT in Sofia (Bulgaria)

Prof. Dr. Eneko Agirre, HiTZ center, University of the Basque Country UPV/EHU

Prof. Dr. Holger Schwenk, Meta AI

Prof. Dr. Margot Mieskes, University of Applied Sciences Darmstadt

Invited Talks

Claudiu Musat – "Natural Interactions with Foundation Models"

Marco Passarotti – "Life after ERC. What's New and What's Next in the LiLa Knowledge Base"

Iryna Gurevych – "Fact or Fiction: How to Spot and Debunk Misleading Content?"

Eneko Agirre – "LLMs and low-resource languages"

Holger Schwenk – "Beyond token-based Large Language Models"

Margot Mieskes – "R^3 - Responsible - Replicable - Research"

Table of Contents

1 Scientific Track **1**

20min-XD: A Comparable Corpus of Swiss News Articles 2
Michelle Wastl, Jannis Vamvas, Selena Calleri and Rico Sennrich

Assessing Open-Weight Large Language Models on Argumentation Mining Subtasks 12
Mohammad Yeghaneh Abkenar, Weixing Wang, Hendrik Graupner and Manfred Stede

Detecting Greenwashing in ESG Reports: A Comparative Analysis of Machine Learning Methods in Traffic-Related Emissions Disclosure 26
Johannes Florstedt, Jonas Fahlbusch and Moritz Sontheimer

Enhancing Multilingual LLM Pretraining with Model-Based Data Selection 32
Bettina Messmer, Vinko Sabolčec and Martin Jaggi

Fine-tuning Whisper on Low-Resource Languages for Real-World Applications 58
Vincenzo Timmel, Claudio Paonessa, Manfred Vogel, Daniel Perruchoud and Reza Kakooee

GOOSVC: Version Control for Content Creation with Generative AI 67
David Gruenert, Alexandre de Spindler and Volker Dellwo

LLM-based Translation for Latin: Summaries Improve Machine Translation 76
Dominic P. Fischer and Martin Volk

Probing BERT for German Compound Semantics 82
Filip Miletić, Aaron Schmid and Sabine Schulte Im Walde

SLANet-1M: A Lightweight and Efficient Model for Table Recognition with Minimal Computational Cost 90
Nguinwa Mbakop Dimitri Romaric, Andrea Petrucci, Simone Marinai and Jean Hennebert

Simulating Human Interactions for Social Behaviour Coaching 104
Daniela Komenda, Livio Bürgisser, Noémie Käser and Alexandre de Spindler

Soft Skills in the Wild: Challenges in Multilingual Classification 109
Laura Vásquez-Rodríguez, Bertrand Audrin, Samuel Michel, Samuele Galli, Julneth Rogenhofer, Jacopo Negro Cusa and Lonneke van der Plas

Using Phonemes in a Cascaded S2S Translation Pipeline 116
Rene Pilz and Johannes Schneider

embed2discover: the NLP Tool for Human-In-The-Loop, Dictionary-Based Content Analysis . . . 121
Oleg Bakhteev, Luis Salamanca, Laurence Brandenberger and Sophia Schlosser

2 Corpus Track **134**

SwissCoco2025 - The Swiss Corpora Collection 2025 135
Mark Cieliebak, Jonathan Gerber and Manuela Hürlimann

Swiss Parliaments Corpus Reimagined (SPC_R): Enhanced Transcription with RAG-based Correction and Predicted BLEU 151
Vincenzo Timmel, Manfred Vogel, Daniel Perruchoud and Reza Kakooee

SwissGPC v1.0 - The Swiss German Podcasts Corpus 157

Samuel Stucki, Mark Cieliebak and Jan Deriu

3 Impact Track **165**

Abstracts of Impact Track 166

Jonathan Gerber, Mark Cieliebak, Don Tuggener and Manuela Hürlimann