

Tobias Eder

Doctoral Researcher in NLP & Trustworthy AI

✉ tobias@eder.ai 📍 Munich, Germany 🗣️ [airymon](#) 🌐 [tobias-eder-socialtum](#)

Summary

My research focuses on building and evaluating trustworthy NLP systems, particularly in high-stakes and multimodal settings. I investigate how data selection, model introspection, and evaluation design can improve reliability beyond standard metrics and across domains including legal text processing, social media analysis, misinformation detection, and content moderation. I am interested in where current systems fail and what that tells us about the gap between benchmark performance and real-world deployment.

Education

Doctoral Candidate	Technical University of Munich (TUM) , Computer Science <ul style="list-style-type: none">• Topic: Trustworthy NLP — investigating data quality, model robustness, and evaluation design across high-stakes domains.• Supervisor: Prof. Georg Groh, Social Computing Group.• Expected graduation: 2026.	Munich, Germany 2021 – 2026
M.Sc.	Technical University of Munich (TUM) , Data Engineering and Analytics <ul style="list-style-type: none">• Thesis: Bias Detection in Hate Speech Datasets.	Munich, Germany 2017 – 2020
B.Sc.	Technical University of Munich (TUM) , Computational Linguistics and Computer Science	Munich, Germany 2015 – 2017
M.A.	Ludwig-Maximilians-Universität München (LMU) , German Literature, Political Science, and Law <ul style="list-style-type: none">• ERASMUS exchange year at Queen Mary University of London (2013–2014).	Munich, Germany 2010 – 2015

Experience

Technical University of Munich (TUM), Social Computing Group , Doctoral Researcher	Munich, Germany 2020 – 2025 5 years
Research, teaching, and project leadership at the intersection of NLP, multimodal learning, and trustworthy AI. Responsible for the full cycle from project acquisition and design through research execution to publication and student mentorship. <ul style="list-style-type: none">• Led the COLEX research project as PI (Software Campus / BMBF, 2022–2024), managing a team of 5. Delivered a working prototype demonstrated at the Software Campus Summit and handed off to industry partner DATEV.• Conducted the technical track of the EMF project (2023–2025): collected and analyzed approximately 400k social media posts across platforms, producing a misinformation landscape analysis that informed the BfS public communication strategy and resulted in a recommendation action plan for misinformation handling.• Contributed to the HUILE project (Siemens collaboration, 2021–2023) on human-in-the-loop techniques for enterprise NLP, delivering an internal prototype handed off to Siemens.• Designed and taught six courses and lab courses across NLP, explainability, and AI ethics (2020–2025).• Supervised approximately 30 master’s theses and 6 bachelor’s theses spanning content moderation, legal NLP, misinformation detection, and explainability.	

Ablacon Inc., Software Developer

Munich, Germany

Oct 2018 – Dec 2019

Full-stack development at a medical AI startup building a mapping system for atrial fibrillation treatment. The company has since been acquired by Boston Scientific.

1 year 3 months

- Built the web frontend and backend (Django, Flask) for Ablacon's Electrographic Flow visualization platform as part of a 6-person technical team.
- Prototype I contributed to entered the clinical trial pipeline for pre-clinical validation.

Siemens Corporate Technology, NLP Intern

Munich, Germany

Jan 2018 – Apr 2018

- NLP and semantic modeling for enterprise applications.

4 months

Research Projects

COLEX: Automatic Analysis and Summarization of German Legal Correspondence

2022 – 2024

Principal Investigator (Software Campus / BMBF / DATEV, 2022–2024). Developed a graph-based retrieval augmentation system leveraging German legal citation structure for improved context in legal document analysis. Created a multi-perspective summarization corpus for court rulings with distinct summary targets. Built an end-to-end prototype with web frontend integrating all components.

- Funded by BMBF via Software Campus.
- Role: Principal Investigator — responsible for project design, hiring, budget, and research direction.

Misinformation Analysis for EMF and 5G Communication

2023 – 2025

Technical lead for the computational track (Bundesamt für Strahlenschutz, 2023–2025). Responsible for social media data collection, quantitative analysis of public discourse on electromagnetic fields, and exploratory work on misinformation detection methods.

- Funded by the German Federal Office for Radiation Protection (BfS).
- Role: Technical lead for data collection and quantitative analysis.

HUILE: Human-in-the-Loop Techniques for Enterprise Use Cases

2021 – 2023

Research on human-model collaboration and feedback-driven learning for enterprise NLP systems (Siemens AG collaboration, 2021–2023).

- Industry collaboration with Siemens AG.

Publications

Safer Reasoning Traces: Measuring and Mitigating Chain-of-Thought Leakage in LLMs

2026

Patrick Ahrend, Tobias Eder, Xiyang Yang, Zhiyi Pan, Georg Groh
(arXiv preprint)

Beyond Hate: Differentiating Uncivil and Intolerant Speech in Multimodal Content Moderation

2026

Nils A. Herrmann, Tobias Eder, Jingyi He, Georg Groh
(CySoc Workshop at ICWSM 2026)

More Than Debunking Misinformation: A Holistic Communication Model for Public Authorities — The Case of Electromagnetic Fields and 5G in Germany

2026

Kerstin Dressel, Larissa Drescher, Katja Aue, Tobias Eder, Georg Groh, Jutta Roosen
(Research Handbook on Public Communication (Edward Elgar), forthcoming)

Retrieving Users' Opinions on Social Media with Multimodal Aspect-Based Sentiment Analysis	2023
Best Paper Award.	
Miriam Anschütz, Tobias Eder, Georg Groh (IEEE 17th International Conference on Semantic Computing (ICSC 2023))	
Comparative Analysis of Cross-Lingual Contextualized Word Embeddings	2022
Hossain Shaikh Saadi, Viktor Hangya, Tobias Eder, Alexander Fraser (2nd Workshop on Multi-lingual Representation Learning (MRL) at EMNLP 2022)	
Long Input Dialogue Summarization with Sketch Supervision for Summarization of Primetime Television Transcripts	2022
Nataliia Kees, Thien Nguyen, Tobias Eder, Georg Groh (Workshop on Automatic Summarization for Creative Writing at COLING 2022)	
Explaining Neural NLP Models for the Joint Analysis of Open- and Closed-Ended Survey Answers	2022
Edoardo Mosca, Katharina Harmann, Tobias Eder, Georg Groh (2nd Workshop on Trustworthy NLP (TrustNLP) at ACL 2022)	
Introducing an Abusive Language Classification Framework for Telegram to Investigate the German Hater Community	2022
Maximilian Wich, Adrian Gorniak, Tobias Eder, Daniel Bartmann, Burak Enes Çakıcı, Georg Groh (Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2022))	
Bias and Comparison Framework for Abusive Language Datasets	2022
Maximilian Wich, Tobias Eder, Hala Al Kuwatly, Georg Groh (AI and Ethics, 2(1), 79–101)	
Mediale Hasssprache und technologische Entscheidbarkeit: Zur ethischen Bedeutung subjektiv-perzeptiver Datenannotationen in der Hate Speech Detection	2022
Laurence Lerch, Maximilian Wich, Tobias Eder, Georg Groh (Medien–Demokratie–Bildung (Springer), 295–310)	
Anchor-Based Bilingual Word Embeddings for Low-Resource Languages	2021
Tobias Eder, Viktor Hangya, Alexander Fraser (Proceedings of ACL-IJCNLP 2021 (Volume 2: Short Papers), 227–232)	
A Scenario-Based Approach to the Design and Use of Ethical AI Models in Managing a Health Pandemic	2021
Georg Groh, Dirk Brand, J.V.D. Merwe, M. Hoffmann, Tobias Eder, Edoardo Mosca (Institute for Ethics in Artificial Intelligence, Munich)	
Evaluating Bilingual Word Embeddings on the Long Tail	2018
Fabienne Braune, Viktor Hangya, Tobias Eder, Alexander Fraser (Proceedings of NAACL-HLT 2018 (Volume 2: Short Papers), 188–193)	

Teaching

Technical University of Munich , Lectures	2021 – 2025
<ul style="list-style-type: none"> • Advanced NLP (lecture block on data, bias, and ethics, 2023–2025). • Natural Language Processing (tutorial sessions, WS 2021). 	

Technical University of Munich, Lab Courses 2020 – 2025

- NLP Lab Course (2021–2025): continuous project-based NLP course.
- Ethical AI Lab Course (2022–2025, summer term): tackling ethically challenging problems in ML and AI.
- XAI Lab Course (2020–2023, winter term): explainability methods for machine learning.

Technical University of Munich, Seminars 2021 – 2025

- Advanced NLP Seminar (2022–2025, winter term): recent developments in NLP.
- Ethics for Nerds Seminar (2021–2024, bachelor level): general ethical issues in computer science.
- Ethics in NLP Seminar (2022–2023, summer term): ethical challenges specific to NLP.

Technical University of Munich, Thesis Supervision 2021 – 2025

- Supervised approximately 30 master’s theses and 6 bachelor’s theses across NLP, content moderation, legal text processing, explainability, and misinformation detection.

Professional Service

Peer Review, ACL Rolling Review (since 2023), NeurIPS (2026), AISTATS, ICWSM, and various workshops 2023 – 2026

Professional Membership, Association for Computational Linguistics (ACL) 2021 – 2026

Awards

Best Paper Award 2023

IEEE 17th International Conference on Semantic Computing (ICSC 2023). For: Retrieving Users’ Opinions on Social Media with Multimodal Aspect-Based Sentiment Analysis.

Skills

Programming: Python (primary), C++/Cython, SQL, JavaScript

ML / NLP: PyTorch, scikit-learn, HuggingFace Transformers/Trainer/Accelerate/Evaluate, Sentence-Transformers, spaCy, gensim, LIME, SHAP

LLM & Fine-tuning: PEFT/LoRA, bitsandbytes, vLLM, Ollama, LangChain

Data & Visualization: pandas, NumPy, matplotlib, seaborn, Jupyter

Web & Prototyping: Django, Flask, React, Gradio, Streamlit

Infrastructure: Docker, Git, Linux, RunPod, PostgreSQL, MongoDB, Weights & Biases, TensorBoard, Hydra

Languages: German (native), English (fluent, academic working language)