Jonas Rieger

rieger@statistik.tu-dortmund.de
https://jonasrieger.github.io/

Education

Doctoral degree *[Doktor der Naturwissenschaft]*, Dr. rer. nat. | ***Statistics*** — Sept. 2022
TU Dortmund University — Dortmund, Germany

Master of Science, M. Sc. | ***Statistics*** — Nov. 2018
TU Dortmund University — Dortmund, Germany

Bachelor of Science, B. Sc. | ***Statistics*** — Oct. 2016
TU Dortmund University — Dortmund, Germany

Previous and Current Positions

Scientific Manager of TRR 391 — since Oct. 2024
TU Dortmund University — Dortmund, Germany

Postdoc — since Oct. 2022
TU Dortmund University — Dortmund, Germany

NLP Scientist — Oct. 2022 – Sept. 2023
Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) — Hamburg, Germany

Doctoral student — Dec. 2018 – Sept. 2022
TU Dortmund University — Dortmund, Germany

Affiliations

Member of the TU Dortmund Young Academy — since June 2023

Member of DoCMA (Dortmund Center for Data-based Media Analysis) — since Dec. 2018

Leibniz Institute for Media Research | **Hans-Bredow-Institut (HBI)** — Oct. 2023 – Sept. 2024

Funding

BMWK: Federal Ministry for Economic Affairs and Climate Action — Sept. 2024 – Feb. 2027
Joint project: Social (in)justice in the energy transition — From the digital debate to the living world — 100 875€
Sub-project: Monitoring narratives about the energy transition

RC Trust: Research Center for Trustworthy Data Science and Security — Oct. 2023
Trustworthy performance evaluation of large language models — 21 000€

TU Dortmund Young Academy — June 2023
The era of ChatGPT: Evaluation and regulation of large language models — 5000€

Professional Experience

Preparation and submission of an approved CRC/Transregio grant application — Oct. 2024 – June 2028
TRR 391: Spatio-temporal statistics for the transition of energy and transport — DFG

Projects

Diskurs Energiewende — since Sept. 2024
Monitoring narratives about the energy transition — BMWK

NEAR — since Oct. 2023
Narrative Economics Alliance Ruhr — MERCUR

Further Project Affiliations

Medien-Doktor Assistance since Mar. 2023
KI-Assistenzsysteme für eine bessere Medizinberichterstattung

GADMO since Oct. 2022
German-Austrian Digital Media Observatory EU

FLACA Oct. 2022 – Sept. 2024
Few-Shot Learning for Automated Content Analysis in Communication Science BMBF

Former Projects

GADMO Oct. 2022 – Sept. 2023
German-Austrian Digital Media Observatory EU

FLACA Oct. 2022 – Sept. 2023
Few-Shot Learning for Automated Content Analysis in Communication Science BMBF

Research Interests

Methodological research

- Evaluation of NLP systems (e.g., language models, topic models) in terms of quality, reliability, robustness
- Parameter-efficient fine-tuning (PEFT) for large language models in few-shot scenarios
- Model selection and parameter tuning of topic models
- Update algorithms and monitoring settings for topic models
- Detection of structural breaks, events and narratives in text corpora

Software engineering in R, Python, and Shiny

- Author and maintainer of CRAN packages rollinglda, ldaPrototype, and topiclabels
- Author and maintainer of Python module petapter
- Co-author and contributor of CRAN packages tosca and spINAR
- Co-author and co-project leader of the Medien-Doktor Assistance App, a Shiny app to assist in evaluating incoming and outgoing media articles, currently used by editorial teams at tagesschau, NDR Info and Nürnberger Nachrichten

Applied research

- Text corpus-based indicators, e.g., UPI and IPI
- Content analysis of texts and tweets of political parties and parliamentarians
- Argument mining and extracting narrative elements from news and social media debates
- Characteristics of disinformation and fact-checks

Service to the Research Community

Reviewing

- ACL Rolling Review and ACL-associated workshops (18 papers)
- Advances in Statistical Analysis
- Applied Sciences
- Communication Methods and Measures
- Comparative European Politics
- Computational Intelligence
- Educational and Psychological Measurement
- Electronics
- Frontiers in Artificial Intelligence
- Frontiers in Social Psychology
- PLOS One
- Scientometrics
- Statistical Papers: DOI: 10.1007/s00362-019-01126-7
- Sustainability

Conferences

- Organizing committee of NEAR Conference 2024
- Program committee of workshop ClimateNLP at ACL 2024
- Organizing committee of Statistische Woche 2023

Academic Administration

- Department Board (Fakultätsrat) since July 2024

Invited Talks

Research Colloquium of the Faculty of Business Studies and Economics, JLU — Jan. 2025
Detecting narration changes in economics
utilizing continuous topic modeling and (large) language models — Gießen, Germany

CMStatistics 2024 — Dec. 2024
Monitoring (social) media narratives
combining retrospective few-shot classification with continuous topic modeling — London, England

CompStat 2024 — Aug. 2024
PETapter: A masked-language-modeling classification head
for modular fine-tuning of (large) language models — Gießen, Germany

BDI/BDA-Arbeitskreis Statistik — May 2024
Bekämpfung von Desinformation und Fake-News durch GADMO — Düsseldorf, Germany

ZPID Lecture Series — Dec. 2022
Keep rollin'! The abilities for monitoring growing corpora using RollingLDA — Trier, Germany

Contributions to Conferences and Workshops

DigiKomm & Methoden 2024 — Sept. 2024
Classifying the needle in the haystack? Problemstellungen beim Einsatz von Argument-Mining für kommunikationswissenschaftliche Fragestellungen am Beispiel der Waffenlieferungsdebatte — Hamburg, Germany

DGPuK 2024 — Mar. 2024
Exploring the potential of large language models (such as GPT-4) for
(semi-)automatic content analysis of stances and frames in media texts — Erfurt, Germany

Digital Total — Oct. 2023
Few-shot learning for automated content analysis (FLACA)
in the German media debate on arms deliveries to Ukraine — Hamburg, Germany

DiTox'23 Workshop @LDK 2023 — Sept. 2023
Debunking disinformation with GADMO: A topic modeling analysis
of a comprehensive corpus of German-language fact-checks — Vienna, Austria

Statistische Woche 2023 — Sept. 2023
Bekämpfung von Desinformation durch GADMO: Analyse eines umfassenden
deutschsprachigen Faktencheck-Korpus mithilfe von Topic Modellen — Dortmund, Germany

Statistische Woche 2023 — Sept. 2023
Scrutinizing ChatGPT against few-shot learning with adapter extensions
and XLM-RoBERTa: A case study on identifying claims, arguments and their stance
in the German news media debate on arms deliveries to Ukraine — Dortmund, Germany

ECREA PolComm 2023 — Aug. 2023
Beyond "Master Frames": A semi-automated approach to
studying viewpoint diversity of the media discourse — Berlin, Germany

DGPuK 2023 — May 2023
Few-shot learning for automated content analysis:
Efficient coding of arguments and claims in the debate on arms deliveries to Ukraine — Bremen, Germany

MUFin'23 Workshop @AAAI 2023 — Feb. 2023
Early warning systems? Building time consistent perception indicators
for economic uncertainty and inflation using efficient dynamic modeling — Washington, DC, USA

SDP'22 Workshop @COLING 2022 — Oct. 2022
Finding scientific topics in contionuously growing text corpora — Gyeongju, Republic of Korea

Statistische Woche 2022 — Sept. 2022
Monitoring consistent topics in continuously growing scientific text corpora — Münster, Germany

Text2Story'22 Workshop @ECIR 2022 — Apr. 2022
Dynamic change detection in topics based on rolling LDAs — Stavanger, Norway

DAGStat 2022 Mar. 2022
Improving the reliability of LDA results using LDAPrototype as selection criterion Hamburg, Germany

EMNLP 2021 Nov. 2021
RollingLDA: An update algorithm of latent Dirichlet allocation to construct consistent time series from textual data Punta Cana, Dominican Republic

EDML'20 Workshop @ECML PKDD 2020 Sept. 2020
Assessing the uncertainty of the text generating process using topic models Online

NLDB 2020 June 2020
Improving latent Dirichlet allocation: On reliability of the novel method LDAPrototype Online

Statistische Woche 2019 Sept. 2019
Quantifizierung der Stabilität der Latent Dirichlet Allocation mithilfe von Clustering auf wiederholten Durchläufen Trier, Germany

DGPuK 2019 May 2019
Softwaretools für die Kommunikationsforschung Münster, Germany

DAGStat 2019 Mar. 2019
Measuring stability of replicated LDA runs Munich, Germany

PUBLICATIONS

Dissertation

- Rieger (2022). "Reliability evaluation and an update algorithm for the latent Dirichlet allocation". *TU Dortmund University*. DOI: 10.17877/DE290R-22949.

Peer-reviewed publications

- Loschke, Braungardt, , Rieger, (2025) "What motivates and demotivates energy savings in times of crisis? — An argument mining analysis using X/Twitter data". Energy Efficiency 18(4). DOI: 10.1007/s12053-024-10283-0.
- Rieger, Jentsch, Rahnenführer (2024). "LDAPrototype: A model selection algorithm to improve reliability of latent Dirichlet allocation". PeerJ Computer Science 10.2279. DOI: 10.7717/peerj-cs.2279.
- Lange, Rieger, Jentsch (2024). "Lex2Sent: A bagging approach to unsupervised sentiment analysis". Proceedings of the 20th KONVENS Conference, pp. 281–291. URL: https://aclanthology.org/2024.konvens-main.28/.
- Faymonville, Riffo, Rieger, Jentsch (2024). "spINAR: An R Package for Semiparametric and Parametric Estimation and Bootstrapping of Integer-Valued Autoregressive (INAR) Models". *Journal of Open Source Software* 9.97, p. 5386. DOI: 10.21105/joss.05386.
- Rieger, Yanchenko, Ruckdeschel, von Nordheim, Kleinen-von Königslöw, Wiedemann (2024). "Few-shot learning for automated content analysis: Efficient coding of arguments and claims in the debate on arms deliveries to Ukraine". *Studies in Communication and Media* 13, pp. 72–100. DOI: 10.5771/2192-4007-2024-1-72.
- Krause, Rieger, Flossdorf, Jentsch, Beck (2023). "Visually Analyzing Topic Change Points in Temporal Text Collections". In: *Vision, Modeling, and Visualization*. DOI: 10.2312/vmv.20231231.
- Rieger, Hornig, Flossdorf, Müller, Mündges, Jentsch, Elmer (2023). "Debunking Disinformation with GADMO: A Topic Modeling Analysis of a Comprehensive Corpus of German-language Fact-Checks". In: *Proceedings of the 4th Conference on Language, Data and Knowledge*, pp. 520–531. URL: https://aclanthology.org/2023.ldk-1.56.
- Rieger, Hornig, Schmidt, Müller (2023). "Early Warning Systems? Building Time Consistent Perception Indicators for Economic Uncertainty and Inflation Using Efficient Dynamic Modeling". In: *Proceedings of the 3rd Workshop on Modelling Uncertainty in the Financial World*. URL: https://github.com/JonasRieger/mufin23/blob/master/paper.pdf.
- Bittermann, Rieger (2022). "Finding scientific topics in continuously growing text corpora". In: *Proceedings of the 3rd Workshop on Scholarly Document Processing*, pp. 7–18. URL: https://aclanthology.org/2022.sdp-1.2.
- Lange, Rieger, Benner, Jentsch (2022). "Zeitenwenden: Detecting changes in the German political discourse". In: *Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis*. URL: https://old.gscl.org/en/arbeitskreise/cpss/cpss-2022/workshop-proceedings-2022.
- Rieger, Lange, Flossdorf, Jentsch (2022). "Dynamic change detection in topics based on rolling LDAs". In: *Proceedings of the Text2Story'22 Workshop*. CEUR-WS. URL: http://ceur-ws.org/Vol-3117/.

- Rieger, Jentsch, Rahnenführer (2021). "RollingLDA: An Update Algorithm of Latent Dirichlet Allocation to Construct Consistent Time Series from Textual Data". In: *Findings Proceedings of the 2021 EMNLP-Conference*. ACL, pp. 2337–2347. DOI: 10.18653/v1/2021.findings-emnlp.201.
- von Nordheim, Rieger, Kleinen-von Königslöw (2021). "From the Fringes to the Core - An Analysis of Right-Wing Populists' Linking Practices in Seven EU Parliaments and Switzerland". In: *Digital Journalism*, pp. 1–19. DOI: 10.1080/21670811.2021.1970602.
- von Nordheim, Koppers, Boczek, Rieger, Jentsch, Müller, Rahnenführer (2021). Die Entwicklung von Forschungssoftware als praktische Interdisziplinarität. In: *M&K Medien & Kommunikationswissenschaft* 69, pp. 80–96. DOI: 10.5771/1615-634X-2021-1-80.
- Rieger, Jentsch, Rahnenführer (2020). "Assessing the Uncertainty of the Text Generating Process Using Topic Models". In: *Proceedings of the ECML PKDD 2020 Workshops*. Vol. 1323. CCIS. Springer, pp. 385–396. DOI: 10.1007/978-3-030-65965-3_26.
- Rieger (2020). "ldaPrototype: A method in R to get a Prototype of multiple Latent Dirichlet Allocations". In: *Journal of Open Source Software* 5.51, p. 2181. DOI: 10.21105/joss.02181.
- Rieger, Rahnenführer, Jentsch (2020). "Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype". In: *Natural Language Processing and Information Systems*, NLDB 2020. Vol. 12089. LNCS. Springer, pp. 118–125. DOI: 10.1007/978-3-030-51310-8_11.
- von Nordheim, Rieger (2020). "Im Zerrspiegel des Populismus - Eine computergestützte Analyse der Verlinkungspraxis von Bundestagsabgeordneten auf Twitter". In: *Publizistik* 65, pp. 403–424. DOI: 10.1007/s11616-020-00591-7.

Selected non-peer reviewed publications (preprints, working papers and datasets)

- Rieger, Ruckdeschel, Wiedemann (2024). "PETapter: Leveraging PET-style classification heads for modular few-shot parameter-efficient fine-tuning". DOI: 10.48550/arXiv.2412.04975.
- Shrub, Rieger, Müller, Jentsch (2022). "Text data rule - don't they? A study on the (additional) information of Handelsblatt data for nowcasting German GDP in comparison to established economic indicators". In: *Ruhr Economic Papers* #964. DOI: 10.4419/96973128.
- Jentsch, Mammen, Müller, Rieger, Schötz (2021). "Text mining methods for measuring the coherence of party manifestos for the German federal elections from 1990 to 2021". In: *DoCMA Working Paper* #8. DOI: 10.17877/de290r-22363.
- Rieger, von Nordheim (2021). "corona100d - German-language Twitter dataset of the first 100 days after Chancellor Merkel addressed the coronavirus outbreak on TV". In: *DoCMA Working Paper* #4. DOI: 10.17877/DE290R-21911.

Teaching Experience (at TU Dortmund University, if not specified)

Natural Language Processing | ***Lecture (English)*** — WiSe 2023/24 – 2024/25
Data Mining Cup | ***Seminar (English)*** — SuSe 2019 – 2025
Introduction to Topic Modeling | ***Seminar (English)*** — SuSe 2023
University of Bremen
Text as Data | ***Lecture (English)*** — WiSe 2022/23
Einführung in LaTeX | ***Compact course (German)*** — SuSe 2019 – 2022
Fallstudien I | ***Seminar (German)*** — WiSe 2021/22
Schätzen und Testen | ***Organization (German)*** — WiSe 2021/22
Nichtparametrische Verfahren | ***Exercise (German)*** — WiSe 2020/21
Text Data meets Econometrics | ***Seminar (English)*** — WiSe 2020/21
Entscheidungstheorie - Statistik VI | ***Exercise (German)*** — SuSe 2020
Wahrscheinlichkeitstheorie - Statistik V | ***Exercise (German)*** — WiSe 2019/20
Textdatenanalyse | ***Seminar (German)*** — SuSe 2019

Supervised Theses

Master

- (Nishat Tasnim Ahmed Meem, running)
- Multimodal Time-LLM: Integrating visual and temporal data for time series forecasting using large language models (Yat Chun Fung, running)
- Comparative analysis of products perception using unsupervised machine learning and LLMs (Vipul Chauhan, running)
- Stock recommendation using graph neural networks and economy news data (Lars Grönberg, running)

- Data preparation & training of LLMs in e-commerce settings (Mariia Hrechyn, 2024)
- Exploration of unsupervised language-style transfer (ULST) methods (Rohan Kumar Nayak, 2024)
- Die Zukunft des Sportjournalismus: Einsatz von Sprachmodellen zur Erstellung von Spielberichten (Niklas Herzog, 2024)
- ChatGPT as a negotiator: Analyzing its adherence to principles of proportionality and equality (Veronika Tsishetska, 2024)
- Attaching PET (-like) models to RELATIO to find causal relationships between narratives (Muhammad Mahir Hasan Chowdhury, 2024)
- Semantic shift modelling with graph neural networks (Imene Kolli, 2024)
- Diachronic sense modeling with hierarchical word embeddings (Aymane Hachcham, 2024)
- #FrierenFürDenFrieden: Quantifizierung des Diskurses zum Thema Gas- und Energiesparen auf Twitter durch Anwendung von NLP-Methoden und Textklassifizierung durch RoBERTa (Carmen Loschke, 2023)
- Comparison of diachronic embeddings with pre-trained model embeddings for historical texts (Priyanka Madiraju, 2023)
- Text Data-based nowcasting of German GDP growth using newspaper data (Yuliya Shrub, 2022)
- Resampling strategies for unsupervised sentiment analysis using lexicon-based text embedding methods (Kai-Robin Lange, 2021)

Bachelor

- Können Zeitungsartikel in ökonometrischen Vorhersagen helfen? Eine VAR-basierte Prognose wirtschaftlicher Kennzahlen mithilfe der IPI- und UPI-Indizes (Kjell Noack, running)
- Modellierung der Zielgruppe und Beliebtheit in Deutschland veröffentlichter Manga — Vergleich von Modellen auch mithilfe von Textdatenanalyse (Darya Lukashina, 2024)
- Comparison of active learning techniques for the benefit of data set generation in the field of text mining (Jannik Bloß, 2023)