

Jenseits algorithmischer Filterblasen: Versuch einer automatisierten Presseschau für ausgewogenen Medienkonsum

Yara Döring | Berliner Hochschule für Technik
Felix Biessman | Berliner Hochschule für Technik & Einstein Center Digital Future

Was ist Presspulse?

- Tagesaktuelle Nachrichten Artikel
- Kuratiert durch Redaktion des DLF
- Zuordnung jedes Beitrags zu den 56 Manifesto-Kategorien
- Zeigt politische Domänen, Links-Rechts Einordnung & Medienvergleich



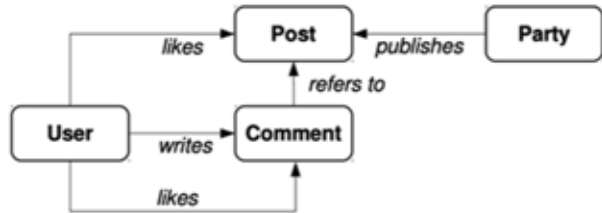
Berliner Hochschule für Technik
Studiere Zukunft



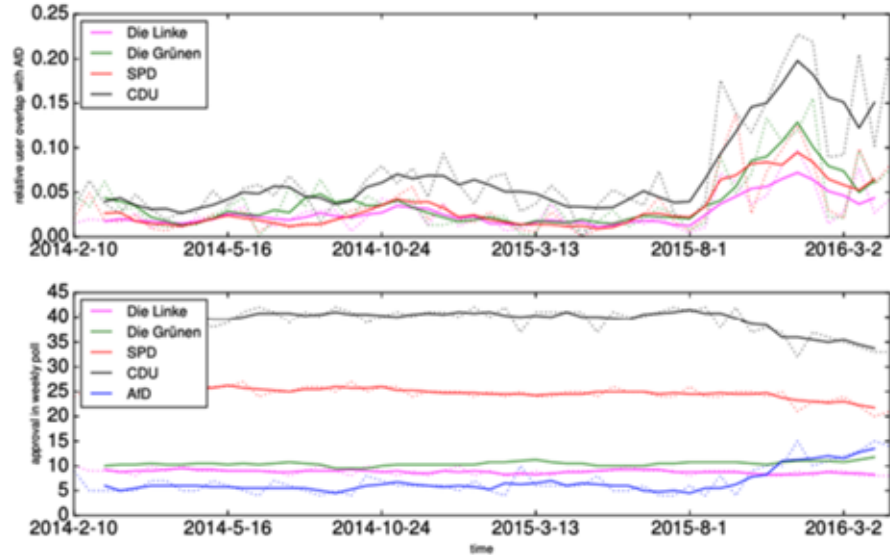
Motivation

- Konsum der meisten Medien über algorithmische Empfehlungen
- Steuerung von Algorithmen durch wirtschaftliche Interessen der Plattformen
- Sie können:
 - politisch einseitigen Medienkonsum verstärken
 - zur Bildung von Echokammern beitragen
- Automatisierte Presseschauen können helfen:
 - unterschiedliche Perspektiven sichtbar machen
 - Erkennung einseitiger Berichterstattung

Beispiel: AfD Nutzerrekrutierung in 2015 auf Facebook



party	posts	comments	likes	users
Die Linke	5,742	74,557	737,437	144,340
Die Grünen	10,031	85,797	581,874	191,022
SPD	1,811	105,234	691,105	111,583
CDU	1,033	133,426	647,348	84,316
AfD	55,586	422,052	3,666,894	306,724
NPD	4,464	378,537	3,791,557	532,999
<i>in aggregate</i>	78,667	1,199,603	10,116,215	1,167,574 ⁴



Schelter et al., Structural Patterns in the Rise of Germany's New Right on Facebook, 2016
 IEEE 16th International Conference on Data Mining Workshops

Probleme Automatisierter Politischer Einordnung

Automatisierte politische Analysen können helfen, algorithmische Biases zu erkennen

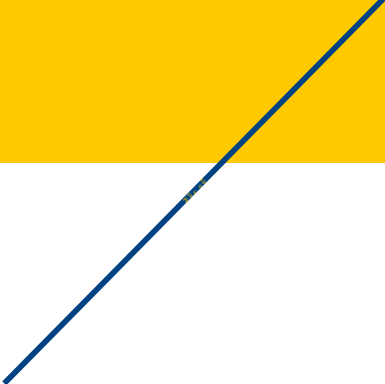
Aber:

- Verfügbarkeit von Trainingsdaten
 - Viele Datensätze nutzen stark vereinfachte Taxonomy (meist binär)
 - Wenig Daten verfügbar für komplexere Taxonomien
- Domain Adaptation
 - Daten aus Parlamentsdebatten / Parteiprogrammen liegen vor
 - Daten, die politische Einordnung benötigen haben andere Sprache

Domain Gap in Automatisierten Politischen Analysen

	Held-out parliament speeches				Party Manifestos			
	precision	recall	f1-score	N	precision	recall	f1-score	N
cducsu	0.62	0.81	0.70	706	0.26	0.58	0.36	2030
fdp	0.70	0.37	0.49	331	0.38	0.28	0.33	2319
gruene	0.59	0.40	0.48	298	0.47	0.20	0.28	3747
linke	0.71	0.61	0.65	338	0.30	0.47	0.37	1701
spd	0.60	0.69	0.65	606	0.26	0.16	0.20	2278
avg / total	0.64	0.63	0.62	2279	0.35	0.31	0.30	12075

Biessmann et al., Predicting political party affiliation from text, International Conference on the Advances in Computational Analysis of Political Text, 2016



Methoden

Benutzte Daten

Trainingsdaten

- Manifesto Project Daten
- Politweets Daten
- Deutsche Zeitungsartikel gesampelt von Fundus

Angewandte Daten

- Deutsche Zeitungsartikel
- Deutsche Twitter Posts

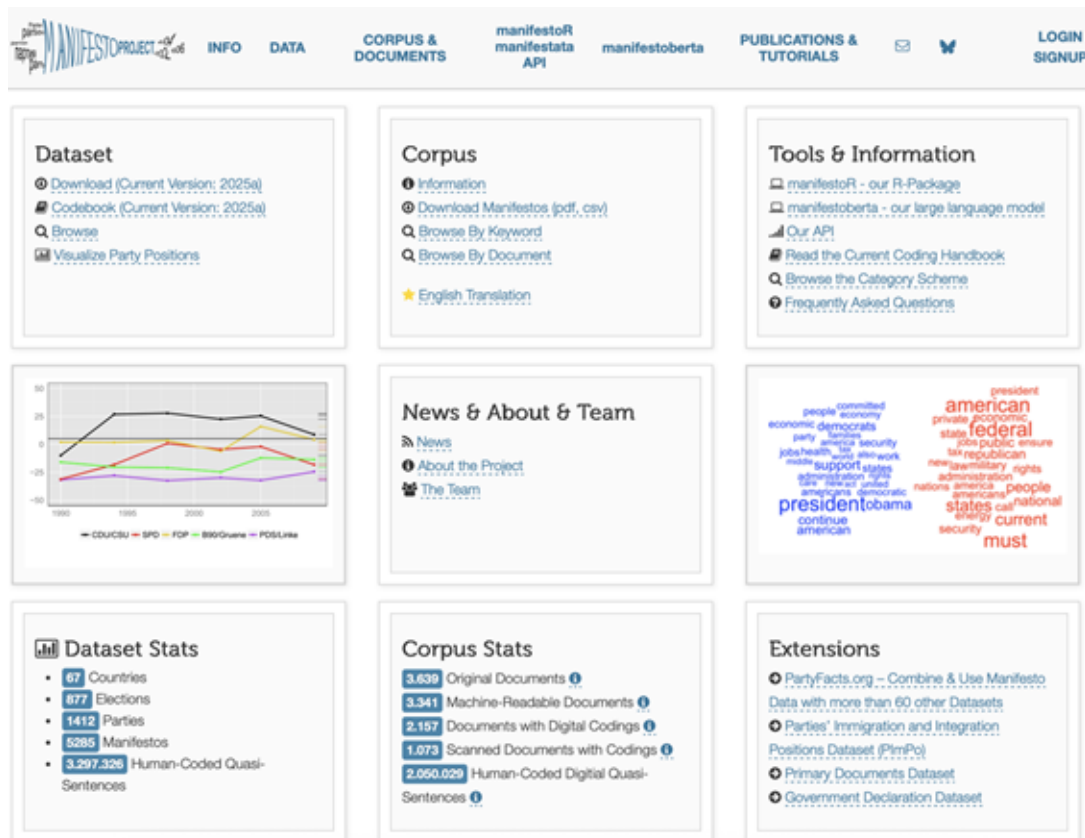
Presspulse

- Presseschau Daten

Manifesto Project



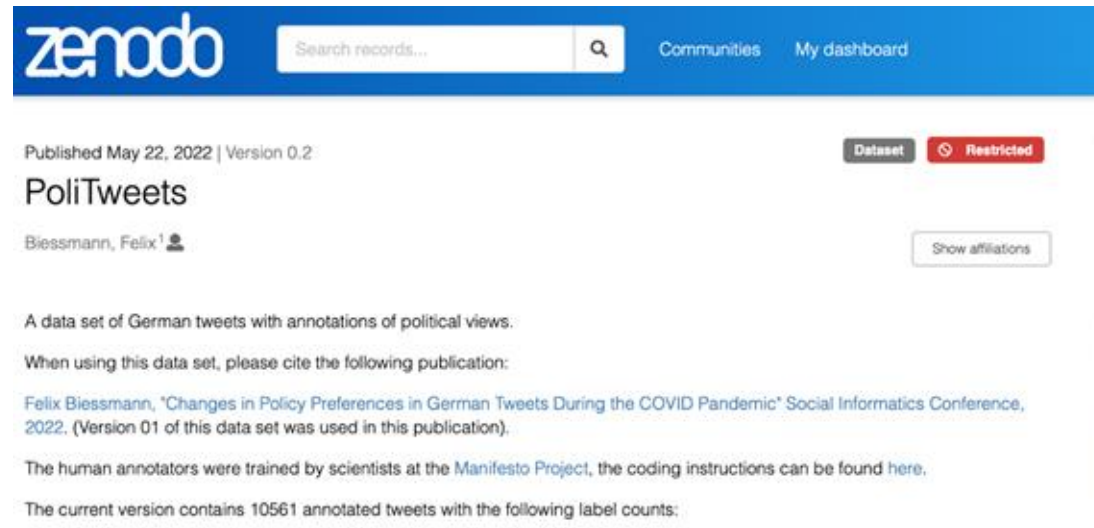
<https://manifesto-project.wzb.eu/>



Manifesto Project: Label Taxonomy

D1 External Relations 10	D2 Freedom and Democracy 4	D3 Political System 5	D4 Economy 16	D5 Welfare and Quality of Life 7	D6 Fabric of Society 8	D7 Social Groups 6
101 Foreign Special Relationships: Positive 102 Foreign Special Relationships: Negative 103 Anti-Imperialism 104 Military: Positive 105 Military: Negative 106 Peace 107 Internationalism: Positive 108 European Community/Union: Positive 109 Internationalism: Negative 110 European Community/Union: Negative	201 Freedom and Human Rights 202 Democracy 203 Constitutionalism: Positive 204 Constitutionalism: Negative	301 Decentralization 302 Centralization 303 Governmental and Administrative Efficiency 304 Political Corruption 305 Political Authority	401 Free Market Economy 402 Incentives: Positive 403 Market Regulation 404 Economic Planning 405 Corporatism/Mixed Economy 406 Protectionism: Positive 407 Protectionism: Negative 408 Economic Goals 409 Keynesian Demand Management 410 Economic Growth: Positive 411 Technology and Infrastructure: Positive 412 Controlled Economy 413 Nationalisation 414 Economic Orthodoxy 415 Marxist Analysis 416 Anti-Growth Economy: Positive	501 Environmental Protection 502 Culture: Positive 503 Equality: Positive 504 Welfare State Expansion 505 Welfare State Limitation 506 Education Expansion 507 Education Limitation	601 National Way of Life: Positive 602 National Way of Life: Negative 603 Traditional Morality: Positive 604 Traditional Morality: Negative 605 Law and Order: Positive 606 Civic Mindedness: Positive 607 Multiculturalism: Positive 608 Multiculturalism: Negative	701 Labour Groups: Positive 702 Labour Groups: Negative 703 Agriculture and Farmers: Positive 704 Middle Class and Professional Groups 705 Underprivileged Minority Groups 706 Non-economic Demographic Groups

Politweets Datensatz



The screenshot shows the Zenodo dataset page for 'PoliTweets'. The header is blue with the Zenodo logo, a search bar, and links for 'Communities' and 'My dashboard'. Below the header, the dataset is titled 'PoliTweets' and is attributed to 'Biessmann, Felix'. It is marked as a 'Dataset' and 'Restricted'. The description states it is a data set of German tweets with annotations of political views. It provides a citation for Felix Biessmann's work at the Social Informatics Conference, 2022. It also mentions that human annotators were trained by scientists at the Manifesto Project and that the current version contains 10561 annotated tweets.

zenodo Search records... Communities My dashboard

Published May 22, 2022 | Version 0.2

PoliTweets

Biessmann, Felix¹

Dataset Restricted

Show affiliations

A data set of German tweets with annotations of political views.

When using this data set, please cite the following publication:

Felix Biessmann, "Changes in Policy Preferences in German Tweets During the COVID Pandemic" Social Informatics Conference, 2022. (Version 01 of this data set was used in this publication).

The human annotators were trained by scientists at the [Manifesto Project](#), the coding instructions can be found [here](#).

The current version contains 10561 annotated tweets with the following label counts:



Presseschau Daten

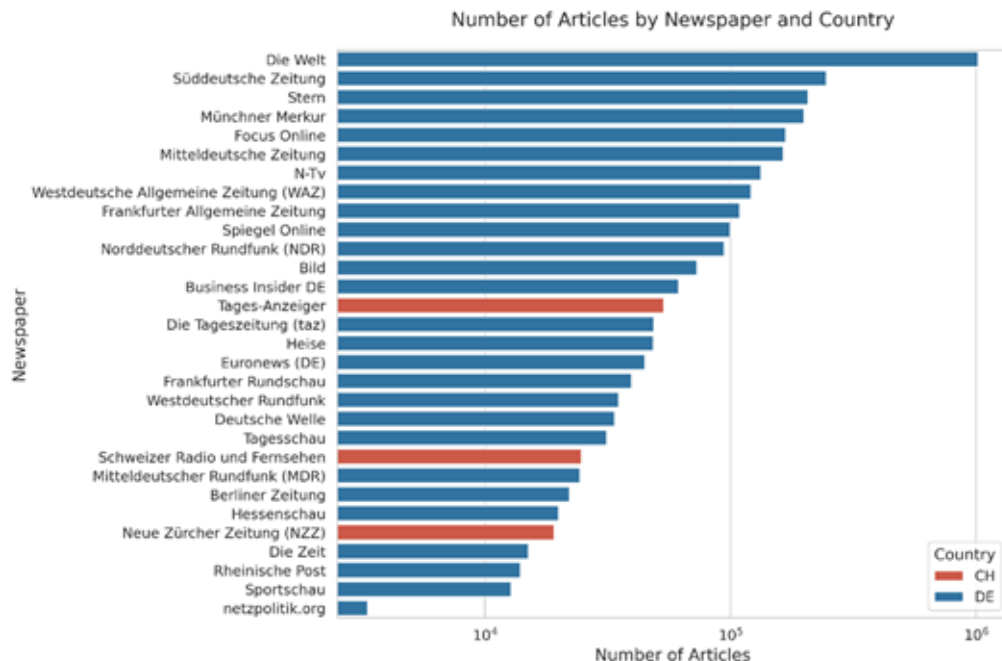
- Tägliches Scrapen der offiziellen Deutschlandfunk-Presseschau
- Extraktion relevanter Inhalte
- Strukturierte Ablage der Daten direkt im GitHub-Repository
- Zentrale Cloud-Datenbasis für nahtlose Nachverfolgung und Historisierung
- Instantane Bereitstellung der aufbereiteten Daten für KI-Analysen und Folgeprozesse



<https://github.com/calgo-lab/presspulse>

Angewandte Daten

- 3.22 Millionen Artikel scraped via Fundus (DE: 3.12 CH: 98K)
- 331.819 politische Tweets gefiltert via Manifesto classifiers

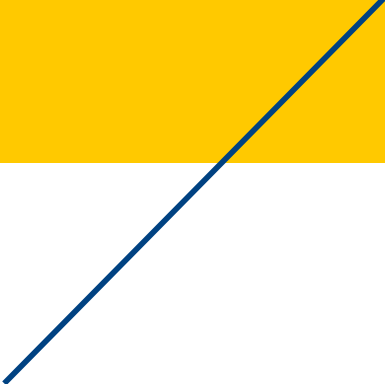


Model Pipeline

- **Architecture:** XLM-RoBERTa / ManifestoBerta finegetuned auf Experten-annotierten Nachrichtenartikeln und Tweets.
- **Ausführung:** Texte wurden in Abschnitte (Chunks) von 512 Tokens unterteilt und die Wahrscheinlichkeiten der Abschnitte pro Artikel aggregiert.
- **Metriken:** Evaluierung über 56 Kategorien mit wettbewerbsfähigen makro-gemittelten
- **F1** : 0.58, **Precision:** 0.58, **Recall:** 0.60

Related work

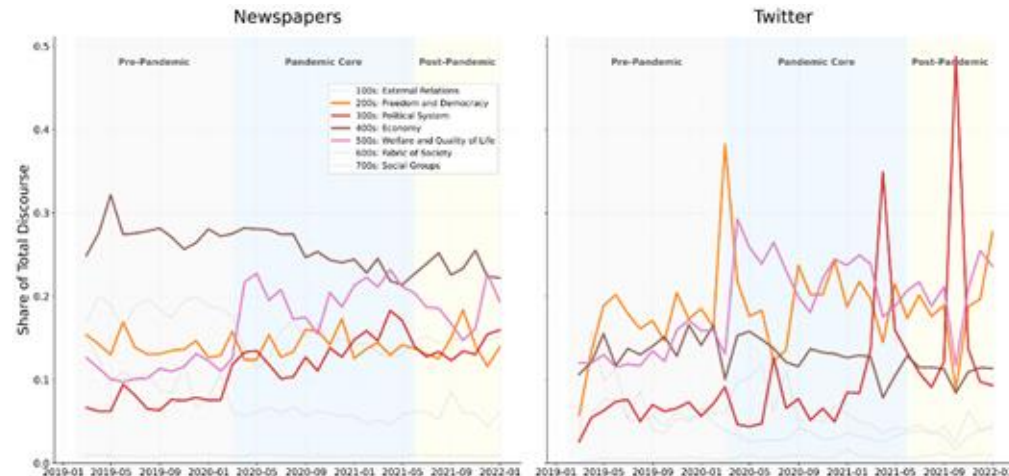
- **SemEval Task 4: Hyperpartisan News Detection (Kiesel et al., 2019):** Demonstriert die Machbarkeit von Transformer-Modellen (Genauigkeit: 0,822) bei der binären Klassifikation politischer Verzerrungen (Political Bias).
- **Social Media and Policy Responses to COVID-19 (Gilardi et al., 2021):** Zeigt ein bi-direktionales plattformübergreifendes Agenda-Setting, wobei Social Media die Debatten in den klassischen Medien anführt.
- **The Manifesto Project (MARPOR) (Lehmann et al., 2025):** Liefert ein 56 Kategorien umfassendes Kodierschema über 7 Politikfelder hinweg für feingliedrige Analysen.
- **ManifestoBerta & German COVID Tweets (Burst et al., 2023, Bießmann, 2023):** Offenbart Grenzen beim Domaintransfer, wenn Manifesto-Modelle auf Nachrichtenartikel oder Twitter angewendet werden



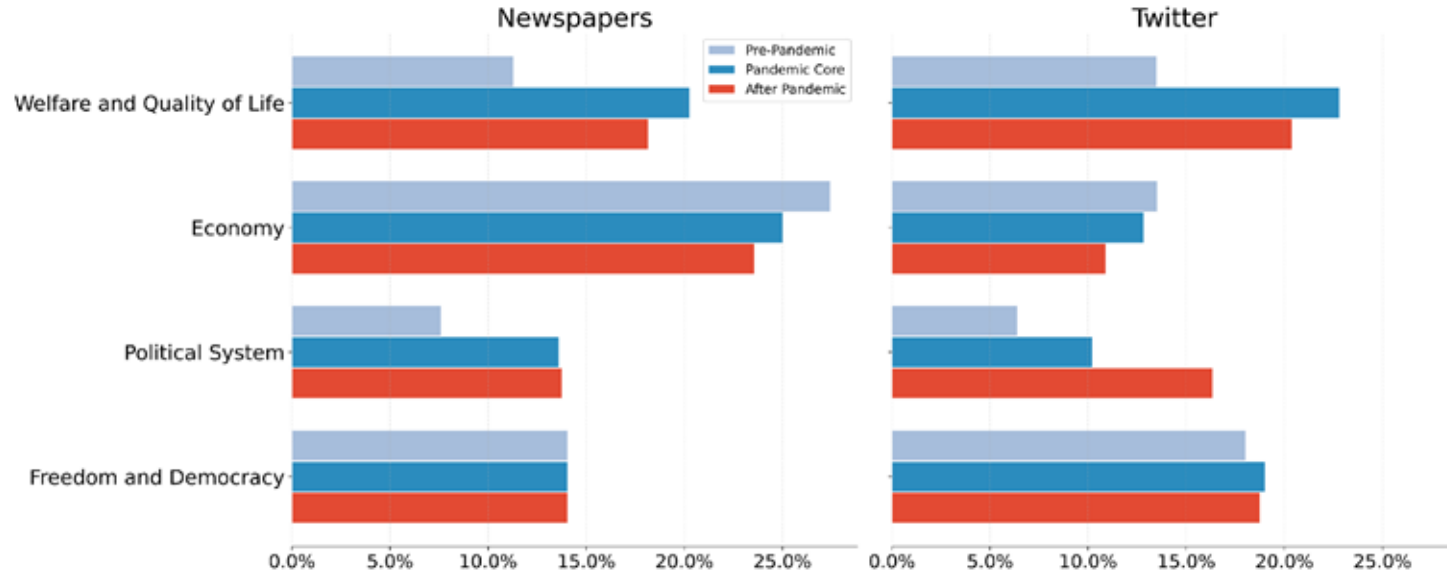
Ergebnis

Politische Themen in Nachrichten und auf Twitter während COVID-19

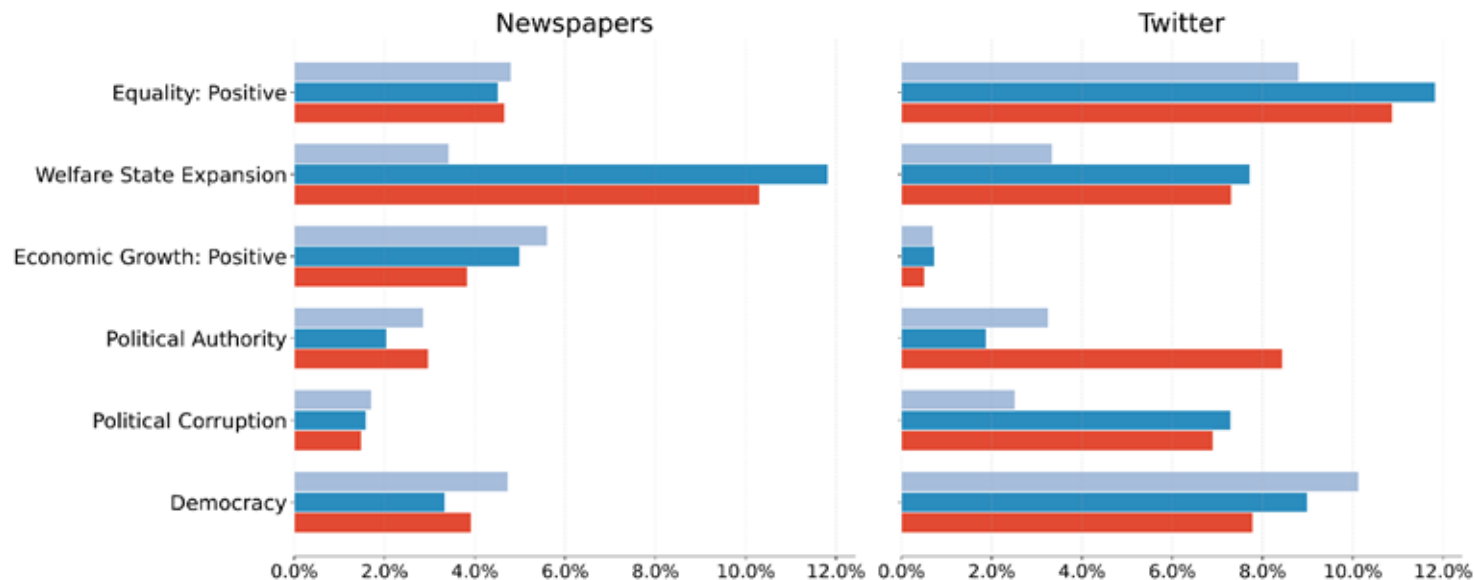
- **Pre-Pandemic (2019):** Zeitungen werden strukturell vom Thema *Economy* dominiert (25–30 %), während Twitter diffuse, unstrukturierte Themen aufweist
- **Pandemic Core (2020–2021):** Beide konvergieren rund um das Thema *Welfare and Quality of Life*
- **Post-Pandemic (Spät 2021+):** Twitter zeigt kurzzeitige, extreme Ausschläge beim Thema *Freedom & Democracy*



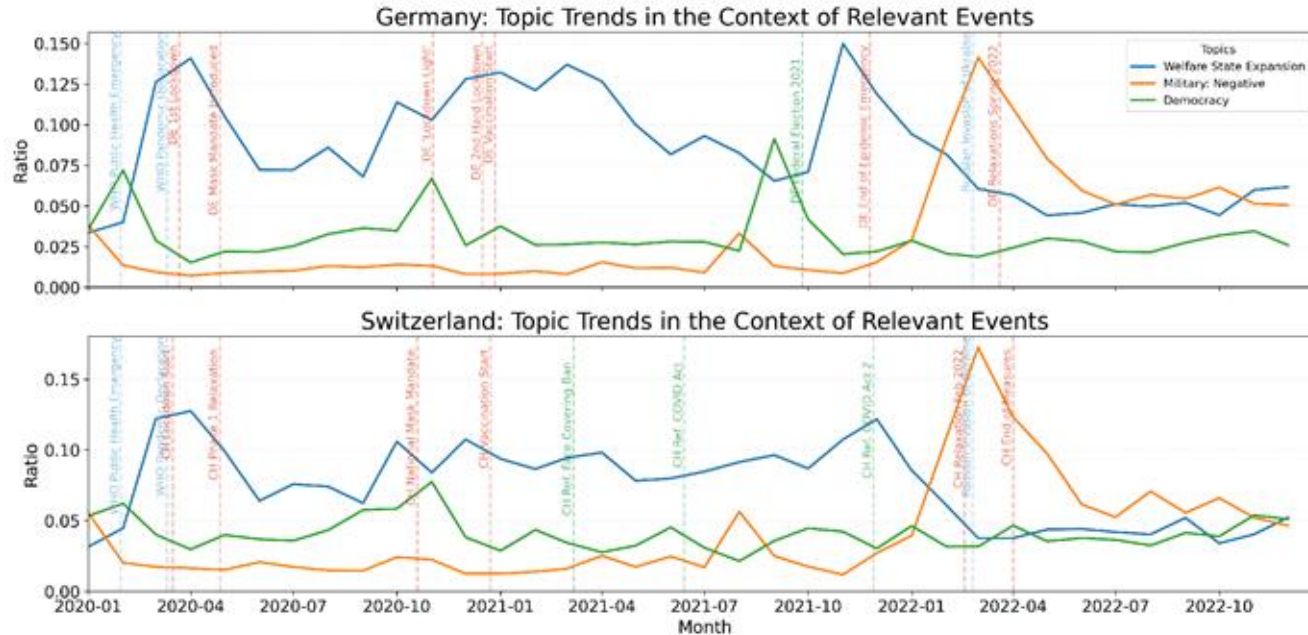
Domain



Label



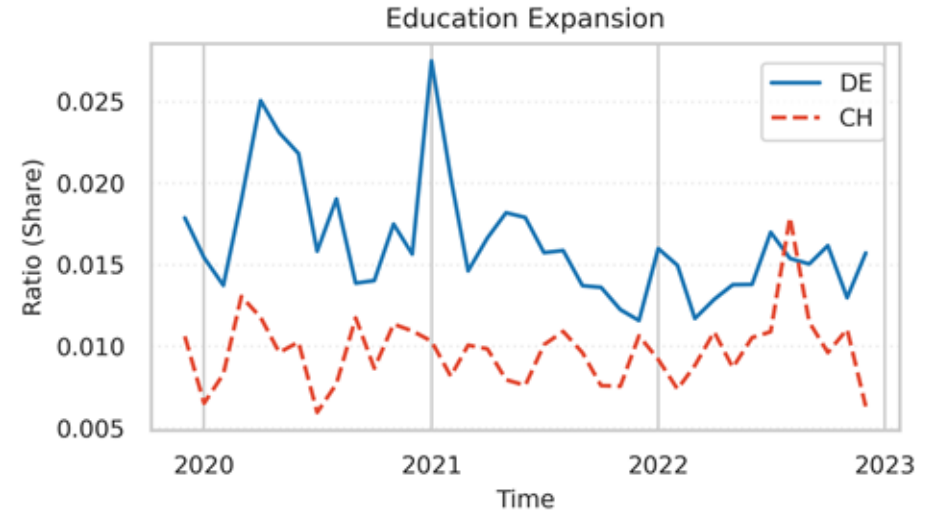
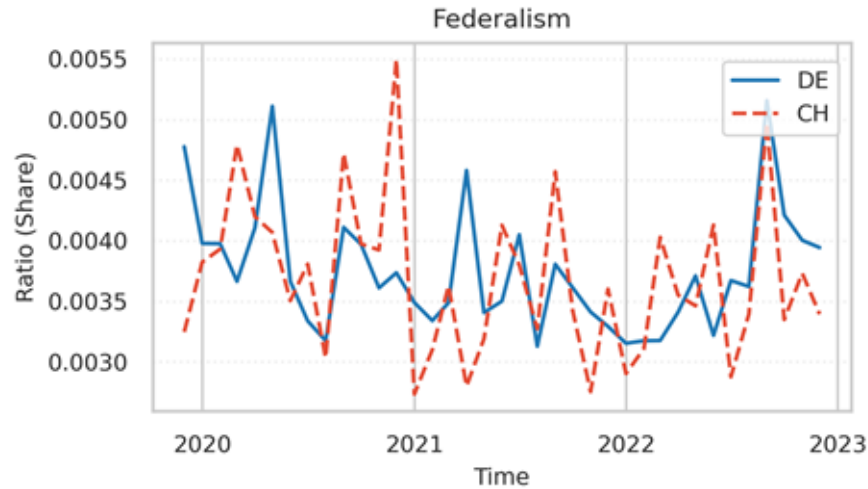
Zeitliche Dynamik in Schweizer und deutschen Nachrichten:



Zeitliche Dynamik in Schweizer und deutschen Nachrichten:

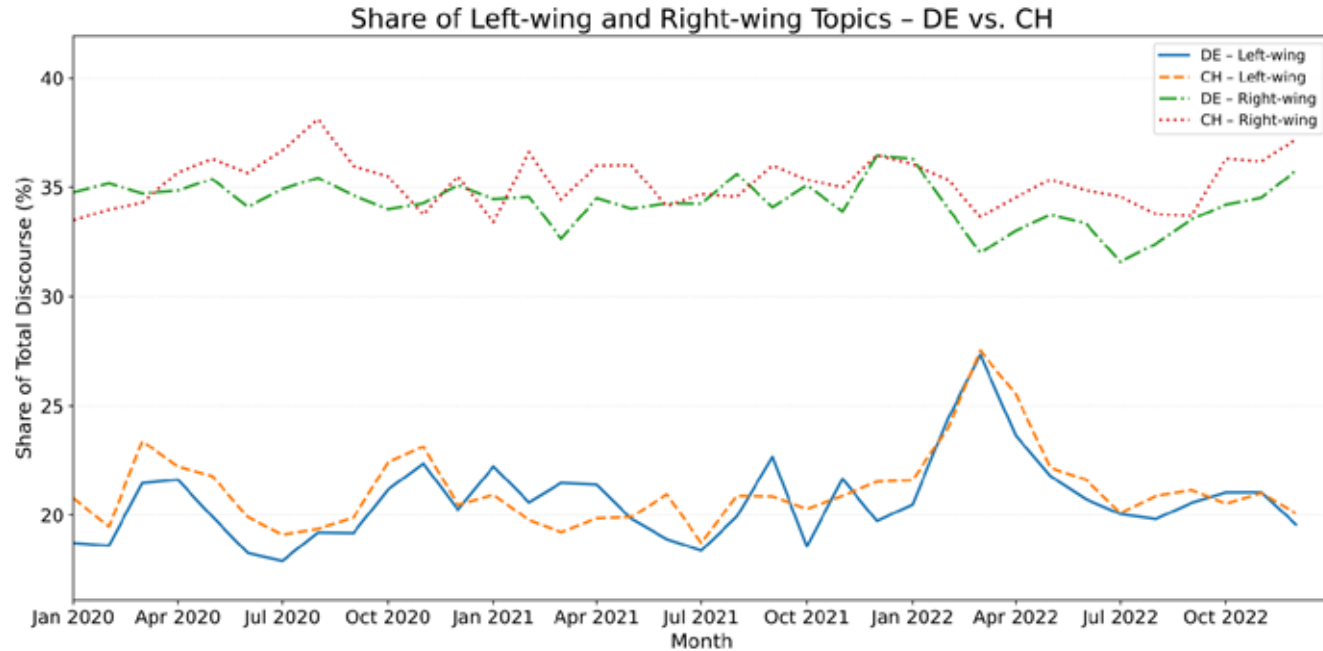
- Mehrheit der 56 Manifesto-Kategorien zeigt positive Korrelationen über die Ländergrenzen hinweg
- Themen mit hoher Konvergenz:
 - **Military: Negative ($p=0.93$):** Synchroner, steiler Anstieg im Februar 2022 infolge des russischen Angriffskriegs auf die Ukraine
 - **Welfare State Expansion ($p=0.89$):** Parallele Spitzenwerte während der ersten COVID-19-Lockdown-Maßnahmen im März 2020
 - **Anti-Imperialism ($p=0.85$):** Paralleles, international ausgerichtetes Framing.

Themen mit niedriger Konvergenz



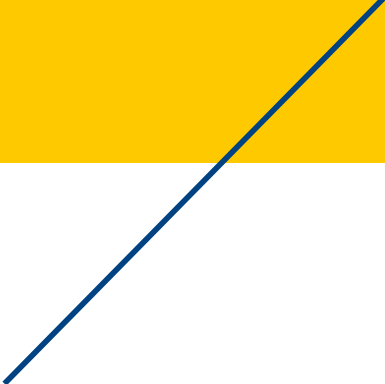
Unterschiedliche institutionelle Ansätze bei Schul-Lockdowns & bei zentraler vs kantonaler Entscheidungsfindung

Vereinfachung politischer Präferenzen auf eine Links-Rechts-Orientierung:



Vereinfachung politischer Präferenzen auf eine Links-Rechts-Orientierung:

- **Right-Coded Topics (~35%):** Durchgehend höherer Anteil am gesamt Diskurs, getrieben durch eine starke Berichterstattung über Konstitutionalismus, Freiheit und Wirtschaftswachstum
- **Left-Coded Topics (~20%):** Stabile Basislinie mit einem gemeinsamen Ausschlag Anfang 2022 (*Military: Negative*)
- Die Schweiz weist einen um 3–4 % höheren Anteil an rechts-kodierten Themen auf, was die Repräsentation marktliberaler Medienhäuser (wie NZZ oder Tages-Anzeiger) widerspiegelt



Fazit

Limitations - Sampling Biases

- Der deutsche Nachrichten-Datensatz (3,12 Mio.) übersteigt die Schweizer Stichprobe (97,9k) bei Weitem
- Einzelne große Medienhäuser beeinflussen die Gesamtzahlen stark
- Historische Tweets hängen von den Keyword-Filter-Protokollen von Drittanbietern ab

Fazit

- Globale Krisen synchronisieren Themen über Länder- und Mediengrenzen
- **Zeitungen**: stabiler Anker im öffentlichen Diskurs.
- **Twitter**: ereignisgetriebener Verstärker.



Yara Döring, Felix Bießmann, Global Crises and National Policies: A Large Scale Analysis of Political Content in German Language Online Media, Proceedings of the 2nd International Workshop on Sustainable Artificial Intelligence for addressing Online Information Disorder (SAI4OID) at ECML-PKDD

Presspulse ausprobieren



<https://calgo-lab.de/presspulse/>



**Danke für eure
Aufmerksamkeit!**

BHT

Berliner Hochschule
für Technik

Studiere
Zukunft