Debunking Disinformation with GADMO: A Topic Modeling Analysis of a Comprehensive Corpus of German-language Fact-Checks

Jonas Rieger[†] and **Nico Hornig**[‡] and **Jonathan Flossdorf**[†] and **Henrik Müller**[‡] and **Stephan Mündges**[‡] and **Carsten Jentsch**[†] and **Jörg Rahnenführer**[†] and **Christina Elmer**[‡]

[†]Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany
{rieger,flossdorf,jentsch,rahnenfuehrer}@statistik.tu-dortmund.de
[‡]Institute of Journalism, TU Dortmund University, 44221 Dortmund, Germany
{nico.hornig,henrik.mueller,stephan.muendges,christina.elmer}@tu-dortmund.de

Abstract

In the age of (semi-) automated creation, reproduction and dissemination of misinformation, manual fact-checking can be considered as a relevant pillar of democracies. To examine the selection mechanisms of fact-checking units, the fact-checks provide a valid basis. Thus, many analyses in the field of natural language processing (NLP) regarding the spread of misinformation are based on the evaluation of fact-checks. We analyze a large German-language fact-check corpus from four specialized newsrooms over the last five years and provide scripts to reproduce the corpus and essential preprocessing steps needed to ensure comparability over time. Our topic model analysis utilizing LDA reveals a strong correlation between current events like Covid and the topics covered by fact-checks in that time. It also shows striking patterns between claims on specific topics and the ratings given by the fact-checkers. In addition, we can show that all considered fact-checking organizations focus primarily on Facebook as a source for the claims they investigate. Cross-cutting topics such as image/video analysis and data-focused fact-checking remain consistent throughout the period.

1 Introduction

In times of dynamic digital publics with significant impacts on reality, quality media cannot ignore the phenomenon of disinformation. Deliberately spreading misinformation poisons public discourse spaces (Lewandowsky et al., 2020) and undermines trust in journalistic actors and institutions by discrediting them or questioning their methods through fabricated arguments (Ognyanova et al., 2020; Giglietto et al., 2019). To counter these negative effects, specific routines and formats have developed in journalism. Probably the best known is the fact-check, in which claims are examined for their degree of truth based on often extensive investigations (Li et al., 2022).

Due to their widespread distribution and the mostly difficult access to often incoherent platform data, it is difficult to examine disinformation campaigns in a comprehensive manner (Bastos, 2022). While, to a certain extent, the topics of the published fact-checks can be used as a proxy variable (cf., Vosoughi et al., 2018) to assess relevant disinformation campaigns, it should be taken into account that the contents of fact-checks may also reflect the media's topic selection criteria, their working routines as well as prevailing trend topics. Consequently, a derivation to the field of disinformation campaigns can only be made to a limited extent.

In this paper, we aim to gain a deeper understanding of the topics covered by fact-checkers in Germany and Austria and their selection mechanisms with regard to the topics and origins of the claims investigated. Therefore, we built, preprocessed, analyzed and provide an extensive German-language fact-check corpus including publications from the past five years from four newsrooms specialized in this beat. The underlying research was made possible by a collaboration within the German-Austrian Digital Media Observatory (GADMO), a cooperation of fact-checkers and scientists co-funded by the European Union, see Section 1.2 for more details and related efforts.

The results show a strong relation of the fact-checks to current events — especially those with a potential for politically motivated campaigns. Clearly assignable switches in the priority topics also point to the limited resources of the news-

Equal contribution.

rooms, as well as attention-economy effects. In addition, all fact-checking organizations focus, with varying degree, on facebook as a source for claims investigated. Cross-cutting themes, on the other hand, appear consistently throughout the period studied — for example, research on images and videos or the focus on data and figures in the fact-checks.

1.1 Related work

In the last three years, the fear of disinformation in Germany has increasingly risen (Hirndorf and Roose, 2023). Whereas in a 2021 survey around 56% indicated that they had great or very great fear, in 2023 this proportion rose to 64%. At the same time, media confidence has declined continuously over the past 8 years (Austria: 48% in 2015 → 41% in 2022, Germany: 60% in 2015 → 50% in 2022), meanwhile at least stagnating again for a few years (Newman et al., 2022).

Along with greater public awareness of the problem of disinformation, the number of fact-checking organizations worldwide has increased in recent years (Amazeen, 2020). While the Duke Reporters' Lab, which maintains a database of fact-checking organizations worldwide, counted 113 such organizations in 2016 (Graves and Cherubini, 2016), it lists 391 active groups as of May 2023[1], ten of which are located in Germany and Austria. However, the effectiveness of fact-checking in countering the belief in disinformation has been widely debated. In some cases, this has led to the conclusion that debunking has no significant effect on reducing belief in disinformation (Schwaiger, 2022). Meta-studies show that fact-checking generally has a positive effect in correcting political disinformation (Walter et al., 2020). It should be noted, however, that the effect is moderated by pre-existing beliefs, ideology and knowledge, and that the evidence on the effect on behavior and knowledge is equivocal (Ecker et al., 2022).

In addition to research on the effectiveness of fact-checking, another body of literature has focused on fact-checkers, their motivations, principles, and purposes, but "virtually no research has conducted a systematic content analysis of fact-checking" (Kim et al., 2022, p. 781). Blum (2020) therefore asks: "Who checks the fact-checkers?" (translated from German). One exception is Humprecht (2020), who analyzes a sample of eight fact-checkers from the United States, the United Kingdom, Austria and Germany with regard to the degree of source transparency provided. She finds that source transparency varies according to the level of journalistic professionalism and organizational differences. However, she uses manual quantitative content analysis, which allows for a more precise understanding of individual texts, but limits the number of observations that can be analyzed.

Automated content analysis, which enables the viewing of a larger number of texts, is used more frequently for viewing disinformation. With regard to the methodological evaluation of alternative media, topic models, such as the latent Dirichlet allocation (LDA, Blei et al., 2003), are often used. For example, Müller and Freudenthaler (2022) analyze a selection of semi-professional German language alternative media using LDA. They show that between 45% and 50% of the content is related to right-wing or populist politics. von Nordheim et al. (2021) were able to show that right-wing populist parties in countries with high media trust tend to share links with a lower source insularity if they are integrated into the party landscape (e.g., Austria), while non-integrated parties (e.g., AfD in Germany) rely more heavily on (their own) alternative media. For both type of parties, the authors were able to detect a high level of thematic insularity by using LDA.

1.2 GADMO

The basis of this study is a project funded by the European Union on combating disinformation. The German-Austrian Digital Media Observatory (GADMO) began its work at the end of 2022 and is the largest alliance of fact-checkers and academic researchers in Germany and Austria. For the first time, the leading fact-checking organizations in Germany and Austria are collaborating closely: the German Press Agency (dpa), the international news agency Agence France-Presse (AFP), the Austrian Press Agency (APA) and the non-profit independent newsroom CORRECTIV. Their work forms the core of the project and is constantly being published on the GADMO website as a new central platform for fact-checks in German[2].

The objectives of the GADMO project also in-

[1] `https://reporterslab.org/fact-checking/`

[2] `https://gadmo.eu/en/gadmo-online-platform-launched/`

clude fostering media literacy, monitoring the platforms regarding overarching policies[3] and researching the field of disinformation. The latter is addressed by two project partners: The Austrian Institute of Technology explores ways in which AI-driven systems can assist journalists to identify manipulated multimedia contents. The team at TU Dortmund University is dedicated to research at the interface between media and data science: On the one hand, the team is interested in fact-checkers, their selection processes, what they cover compared to traditional media and how this differs between different organizations. Therefore, we provide and analyze the German-language fact-check corpus presented in this paper. On the other hand, further work will use network analysis to investigate whether disinformation campaigns can be identified through targeted dissemination patterns[4].

Being part of the European Digital Media Observatory (EDMO), GADMO is integrated into a Europe-wide network of media and research affiliates[5]. In addition, there are close links to projects funded in the Federal Government's research framework program on IT security, which are also intended to counteract the massive spread of disinformation[6]. In this context, the noFAKE[7] project, also aiming at developing an assistance system for the early detection of false information, is particularly worth mentioning.

1.3 Contribution

Our contribution to research is threefold: First, we provide a corpus of about 5000 German-language fact-checks that is reproducible and extensible, thus enabling researchers to carry out further (content) analyses. This is important, as outlined in Section 1, because there is a lack of research on the texts of fact-checks and their characteristics, such as sources and topic decisions. Second, during our data collection process we identified issues such as missing (meta) data or poor comparability between different fact-checking organizations, for which we provide solutions how to address these. Third, we give insights into the topics being considered, the ratings being given, the sources of the claims being investigated and how these differ between different fact-checking organizations.

[3] https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

[4] https://gadmo.eu/en/research-development/

[5] https://edmo.eu/edmo-at-a-glance/

[6] https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/2022/02/fake-news-bekaempfen.html

[7] https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/nofake

2 Data

Our corpus consists of data from four German-language fact-checking organizations: The German language service of Agence France-Presse (AFP), the Austrian Press Agency (APA), the non-profit newsroom CORRECTIV and the German Press Agency (dpa). In the following, we provide a brief overview of the data collecting process. All scraping and analysis scripts are available under `https://github.com/GADMO-EU/DiTox2023`.

2.1 Composition

We allocated the data in a three-step approach: As a starting point for data acquisition, we used the R (R Core Team, 2023) package httr (Wickham, 2022) to access a Google API referencing *ClaimReview*[8], a tagging system that provides fact-check results and their metadata such as publication date, source, and claim rating in a structured way. In a next step, we scraped the texts corresponding to the metadata directly from the respective websites using the R package rvest (Wickham, 2021). As the dpa stopped using ClaimReview in July 2020 when it changed its publication platform, we also scraped the available metadata (publication date and claim). In a third step, we compared the resulting corpus with data provided by the fact-checking organizations as part of our GADMO collaboration. Finally, we restricted the corpus to fact-checks until the end of January 2023.

2.2 Cleaning

Due to the heterogeneous nature of the corpus, some cleaning was necessary. First, we removed duplicate texts identified by the same URL or the same text. In some cases, especially for fact-checks authored by CORRECTIV, we kept very similar texts if they refer to different URLs. As the dpa did not use ClaimReview throughout the whole analysis period, we identified the URL of the analyzed claim manually for most of the data. The same applies to some of the other organizations' fact-checks. In some cases, e.g., when fact-checkers have debunked a phenomenon that was widespread on social media, they did not provide a specific

[8] https://schema.org/ClaimReview

	AFP			APA			CORRECTIV			dpa		
Period	$\lvert D\rvert$	$\lvert W\rvert$	$\bar{N}$	$\lvert D\rvert$	$\lvert W\rvert$	$\bar{N}$	$\lvert D\rvert$	$\lvert W\rvert$	$\bar{N}$	$\lvert D\rvert$	$\lvert W\rvert$	$\bar{N}$
2018/1	·	·	·	·	·	·	132	23 983	182	·	·	·
2018/2	·	·	·	·	·	·	151	34 382	228	·	·	·
2019/1	·	·	·	·	·	·	147	33 318	227	20	3281	164
2019/2	·	·	·	·	·	·	190	58 297	307	211	31 996	152
2020/1	·	·	·	40	10 155	254	223	75 839	340	179	33 294	186
2020/2	86	35 875	417	59	17 677	300	215	84 335	392	376	70 069	186
2021/1	191	79 118	414	46	18 914	411	232	81 597	352	300	59 464	198
2021/2	185	93 891	508	36	15 273	424	238	72 155	303	340	64 800	191
2022/1	145	70 726	488	25	8942	358	234	58 087	248	323	65 948	204
2022/2	127	67 169	529	29	9669	333	238	75 604	318	384	76 414	199
2023/1	26	15 311	589	5	1207	241	34	8720	256	62	13 406	216
Total	760	362 090	476	240	81 837	341	2034	606 317	298	2195	418 672	191

Table 1: Number of fact-checks $|D|$, number of words in fact-checks $|W|$ (after all preprocessing steps), and mean number of words per fact-check $\bar{N}$, for the four fact-check organizations per half-year.

URL and therefore left this entry blank. Sometimes more than one URL was mentioned in the text, in which case we decided to consider only the first one mentioned. In contrast, there are fact-checks, in which no specific URL has been mentioned, but the source was given. For these cases, we decided to include the domain, e.g. *facebook.com*, in the dataset.

2.3 Preprocessing

For the later modeling of the texts we applied common preprocessing steps including lowercasing, stopword removal, punctuation removal, number removal, resolving umlauts and tokenization. Then, we kept only those words that contain at least two letters and occur at least five times in the whole dataset, which results in 27 606 vocabularies.

For referencing the set of fact-checks (cf., Section 3.1), we use the notation $D = \{D_m \mid m = 1, \ldots, M\}$, where M denotes the number of all documents. Moreover, $W = \bigcup D_m$ denotes the set of all words.

Figure 1 shows how the total of 5229 fact-checks (with an average of 281 words per document, after preprocessing) are distributed among the four different organizations. Table 1 provides further insight into the distribution of fact-checks and their length over time. It can be seen that all 283 fact-checks from 2018 in our corpus were authored by CORRECTIV. We observed dpa's first fact-checks for June 2019, from APA for February 2020, and from AFP for September 2020. The fact-checks from dpa are on average the shortest with (relatively consistently) 191 words, while AFP uses on average more than twice as many words (476) per fact-check.

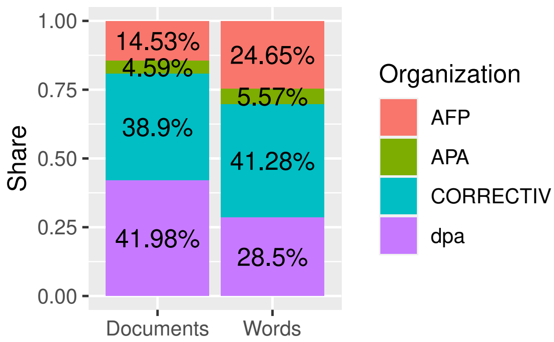

Figure 1: Share of the organizations on the total corpus of fact-checks.

3 Analysis

In the following, we use LDA as a topic model method to automatically present the thematic content from the fact-checks in an unsupervised manner. We also relate the topics identified in this way to the ratings assigned and the sources of the claims examined. Based on the findings from our data analysis we suggest further research questions for future investigations using specialized advanced NLP methods.

3.1 Topic Modeling

To analyze the given dataset, we make use of probabilistic topic modeling, which is used in many application domains (Blei, 2012). In comparison to transformer-based methods (Vaswani et al., 2017), the modeling idea is rather intuitive: a set of documents is described by distributions of topics over time, where each word in each of these documents is assigned to one of the topics. These assignments yield word distributions for each topic, which make the topics interpretable.

Probably the best known topic model is LDA (Blei et al., 2003). The underlying probabilistic model (Griffiths and Steyvers, 2004) can be written as

$$W_n^{(m)} \mid T_n^{(m)}, \phi_k \sim \text{Discr}(\phi_k), \quad \phi_k \sim \text{Dir}(\eta),$$
$$T_n^{(m)} \mid \theta_m \sim \text{Discr}(\theta_m), \quad \theta_m \sim \text{Dir}(\alpha),$$

where α and η are Dirichlet priors for the topic and word distributions, respectively. The number of modeled topics, K, is chosen by the user and each document is considered a bag of words set $D_m = \{W_n^{(m)} \mid n = 1, \dots, N^{(m)}\}$ with observed words $W_n^{(m)} \in \{W_1, \dots, W_V\}$. Then, $T_n^{(m)}$ describes the corresponding topic assignment for each word. Only the words are observable, while all other variables and parameters are latent. The main result, the latent word and topic distributions are represented by ϕ and θ, respectively.

For modeling topics in our German fact-check corpus, we use a reliable variant of classical LDA, estimated with the Gibbs sampler (Griffiths and Steyvers, 2004), named LDAPrototype (Rieger et al., 2022). It selects the medoid LDA — the LDA with the highest mean of pairwise similarities to all other LDAs — from a set of candidate models with independently and randomly initialized topic assignments.

We model all $M = |D| = 5229$ documents together, the vocabulary set is of size $V = 27\,606$. Since Chang et al. (2009) show that the use of common likelihood-based measures, such as perplexity, correlates poorly or even negatively with human perceptions of well partitioned topics, and Hoyle et al. (2021) show that alternative automated measures based on coherence also lead to incoherent decisions, we do not choose automated evaluation measures for parameter tuning. We tried different numbers of topics $5, \dots, 25$ showing $K = 12$ with $\alpha = \eta = 1/K$ to be appropriate in terms of granularity and coherence of topics via human eye-balling.

In the following analysis, we make use of the more reliable medoid LDA (cf., Rieger et al., 2022), which was selected out of 100 independent replications using the R package ldaPrototype (Rieger, 2020).

3.2 Topics

For a better understanding of the automatically generated topics, we let human coders label them. Figure 3 shows the relative frequencies of all $K = 12$ topics in the fact-checks, per organization and overall. Accordingly, *Pictures & Videos* is the most frequently associated topic in AFP fact-checks with 21% of the words assigned to it, while 28% of the words in APA fact-checks are assigned to the topic *Laws & Legal Status*. For CORRECTIV (15% *Corona*) and dpa (12% *Quotes*), the distributions tend to be more balanced, which can to some extent be explained methodologically by the higher number of fact-checks in the analysis, raising the possibility that the smaller subcorpora realize more skewed distributions. From a contents perspective, the connection of AFP fact-checks to image content is plausible since according to their own statements they put a focus on uncovering image manipulation and deep fakes.

One advantage of topic modeling compared to traditional (hard) clustering methods is that the assignment of topics to words, which makes it a soft clustering method, allows, for example, the analysis of co-occurring topics. At the same time, this soft-clustering poses a challenge in determining a precise co-occurrence operationalization. For our analysis, we consider co-occurring topics always in reference to a dominant topic in a particular document. We understand a dominant topic per fact-check as the one that received more than half of all topic assignments in that document. The co-occurrence with other topics can then be computed using the occurrence of all other topic assignments in these associated fact-checks. Using this approach, we obtain the distributions in Figure 2, where *NA* refers to those fact-checks where no dominant topic could be determined.

It can be seen that the topics *Medicine & Health*, *Vaccination* and *Corona* strongly co-occur with each other. For all three (dominant) topics the corresponding two other topics account for about half of the co-occurring assignments. Another observation concerns the topics *Russo-Ukrainian War* and *Pictures & Videos*. While in fact-checks that thematically mainly deal with the war 37% of the remaining words are associated with the topic of im-

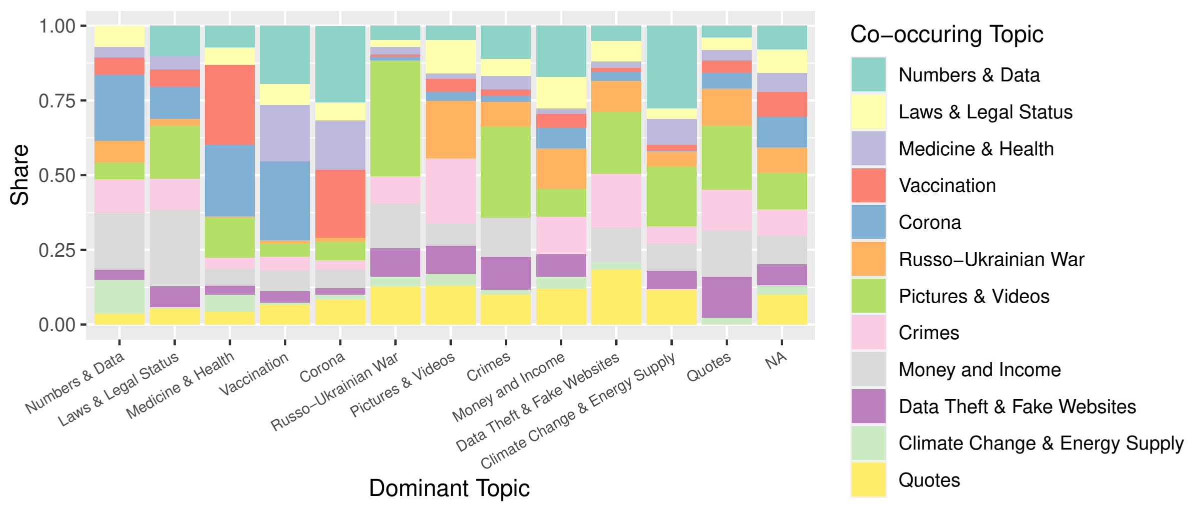

Figure 2: Co-occurring topics in the fact-checks. Dominant topics are considered as those having more than 50% of the topic assignments within the corresponding fact-check. *NA* refers to the absence of a dominant topic for these fact-checks.

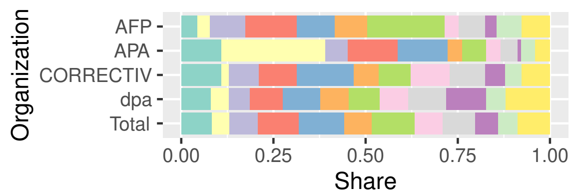

Figure 3: Distribution of the topics in fact-checks of the different organizations; cf., Fig. 2 for legend.

age manipulation, the other way around it is “only” 19%. Furthermore, as a typical side topic, *Pictures & Videos* accounts for 30% of the co-occurrences in *Crimes* fact-checks, for 21% in each of *Quotes* and *Data Theft & Fake Websites*, and for 20% in fact-checks on the topic of *Climate Change & Energy Supply*. The distribution of topics in fact-checks without a dominant topic does not show any particular peculiarities (cf., *Total* bar in Fig. 3).

In addition to the global topic distributions, the changes over time are of special interest. For this purpose, we calculate smoothed values of the number of topic assignments per day and organization using rolling sums over 90 days. To standardize the values, we divide each time series by the maximum of all smoothed values per organization. The intensity of each of the 12 topics over time is shown in Figure 4.

There is a clear focus of CORRECTIV and dpa in particular on Corona-related fact-checks in 2020. Due to the continuously high prevalence of the *Pictures & Videos* topic in AFP fact-checks, this impact is not so clearly visible for their fact-checks. However, the topic *Vaccination* shows a clearly increased prevalence in the second half of 2021, while for APA the topic already becomes more prevalent at the beginning of 2021. The general focus of APA fact-checks on regulations by the state rather than Corona itself is also evident, which in turn explains the high share of this topic *Laws & Legal Status* in Fig. 3. With the start of the war in February 2022, all organizations show a shift in the prioritization of their fact-checks toward the topic *Russo-Ukrainian War*. Overall, the dpa shows the most balanced distribution of topics over the entire period, while the APA shows the clearest focus on one of the modeled topics (cf., Fig. 3).

3.3 Ratings

The analysis of the checked claims’ ratings in the fact-checks is only possible for AFP and CORRECTIV, since APA and dpa do not use a rating scale, but only free-text ratings. Manual review and comparison of the ratings with the textual ratings revealed that there may be occasional incorrect entries. For instance, there was one observation with a rating of 5 and a textual rating of “falsch” (incorrect), while, in general, the AFP fact-checks ratings range from 1-5, with 1 for incorrect and 5 for correct. By correcting this one observation from 5 to 1, AFP fact-checks only realize ratings 1–3 and *NA* (1: 557, 2: 115, 3: 67, *NA*: 21). In Figure 5, the distributions of the ratings in the AFP fact-checks are presented depending on the topic.

According to this, AFP fact-checks assigned to

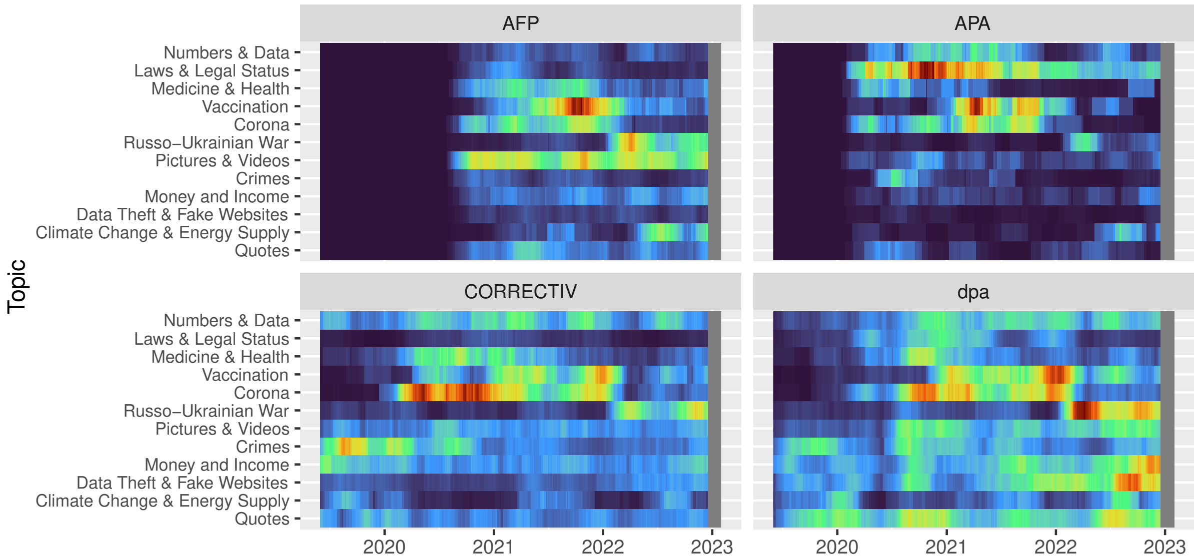

Figure 4: Topic intensity in published fact-checks over time per organization. Values were calculated based on a 90-day rolling window and normalized with the maximum value per organization.

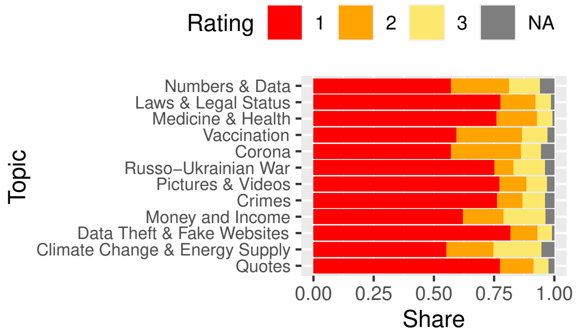

Figure 5: Distribution of the AFP ratings per topic.

the topic *Data Theft & Fake Websites* obtain in over 75% of the cases the lowest possible rating. This topic is thus most often associated with *incorrect* rated claims. Overall, it can be seen that for all topics more than 50% of the correspoding fact-checks obtain rating 1, which can be explained by the global concentration of this rating (73% of the fact-checks). The greatest tendency of a topic to less pronounced degrees of disinformation, i.e., ratings of 2 and 3, can be observed for *Climate Change & Energy Supply*.

In contrast, fact-checks by CORRECTIV are rated on a broader scale of a total of 7 levels identified by us. It is known that CORRECTIV has used a new scale for their rating from October 16, 2020. In this context, the textual ratings *missing context* and *unproved* were added to the scale, which correspond to 4 in the new rating scheme. Table 2 gives the list of textual ratings that occur, their frequencies, and their associated numerical ratings in ClaimReview. The left column in bold reflects the ratings we merged from the old and new schemes.

A manual investigation of individual fact-checks has shown that the numerical rating 2 is also associated with the textual ratings *falscher Kontext* (wrong context) and *manipuliert* (manipulated). Moreover, the ratings *missing context* are also found in fact-checks with the (merged) rating 3, 4, and rarely 6; for all especially for fact-checks before the change of the scheme.

Accordingly, Figure 6 shows that the category *missing context* in light blue has been assigned frequently since its implementation, almost completely replacing *partially incorrect* ratings for some topics. The figure shows the distribution of the ratings over time in relation to the topic. For some topics, the rating 5 temporarily reaches over 50% of the assignments.

A striking pattern is the high number of *NA* values during the Covid pandemic period. We explain this as a result of the inability to check the associated claims conclusively and reliably and because the existing scale did not contain the required rating. With the implementation of rating 5, no more *NA* values occur.

It is notable that assignments to the topic *Data Theft & Fake Websites* occur in up to 50% of cases from fact-checks about claims that are purely fictional. Over time, it also becomes apparent that

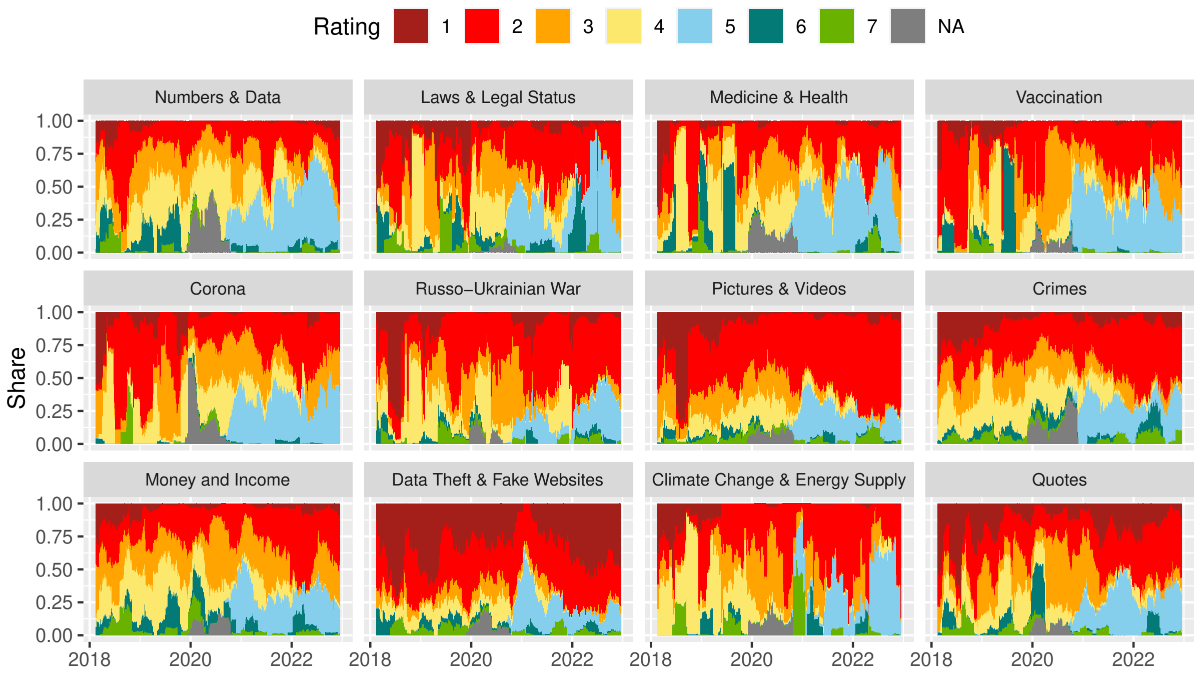

Figure 6: Distribution of the processed merged CORRECTIV ratings (cf., Table 2) per topic over time.

Our	Textual rating	Old	New	$\|D\|$
1	frei erfunden (purely fictional)	1	0	261
2	falsch (incorrect)	2	1	733
3	größtenteils falsch (largely incorrect)	3	2	306
4	teilweise falsch (partially incorrect)	4	3	249
5	fehlender Kontext* (missing context)	·	4	294
6	größtenteils richtig (largely correct)	5	7	75
7	richtig (correct)	6	8	77
NA	·	·	·	54

Table 2: Number of CORRECTIV fact-checks in relation to our processed merged ratings **1** to **7** and **NA**. Until Oct. 15, 2020, an old rating scheme was used, after that a new one. *also includes "unbelegt" (unproved).

Pictures & Videos, beginning in 2021 and probably also due to the co-occurrences in fact-checks on the topic of *Russo-Ukrainian War*, is associated considerably more frequently with false claims from 2022 onward. For the latter topic, we observe an abrupt increase in severe disinformation (ratings 1 & 2) at the beginning of the war.

The topic that is overall less strongly associated with false claims (ratings 1 & 2), but more with misleading claims (3–5) and partly also with correctly rated (6 & 7) claims is *Numbers & Data*. An interpretation is that it seems easy to make a statement with only a few erroneous information or an incorrect integration of percentage, relative or absolute numbers, which either already contains a misinterpretation or consciously accepts this misinterpretation by the reader.

3.4 Domain

We investigated which websites were the source of the claims that were fact-checked. As Table 3 shows, Facebook is the dominant source of claims, accounting for almost 3579 of the 5229 fact-checks in our corpus. This is not surprising, given that three of the four fact-checking organizations examined in this paper cooperate with Meta/Facebook: CORRECTIV since 2017[9], dpa since early 2019, and AFP since 2020. The other 1650 entries are spread across a number of other sites, with only Twitter having more than 200 entries. An *NA* en-

[9] https://correctiv.org/faktencheck/ueber-uns/2018/12/17/ueber-die-kooperation-zwischen-correctiv-faktencheck-und-facebook/

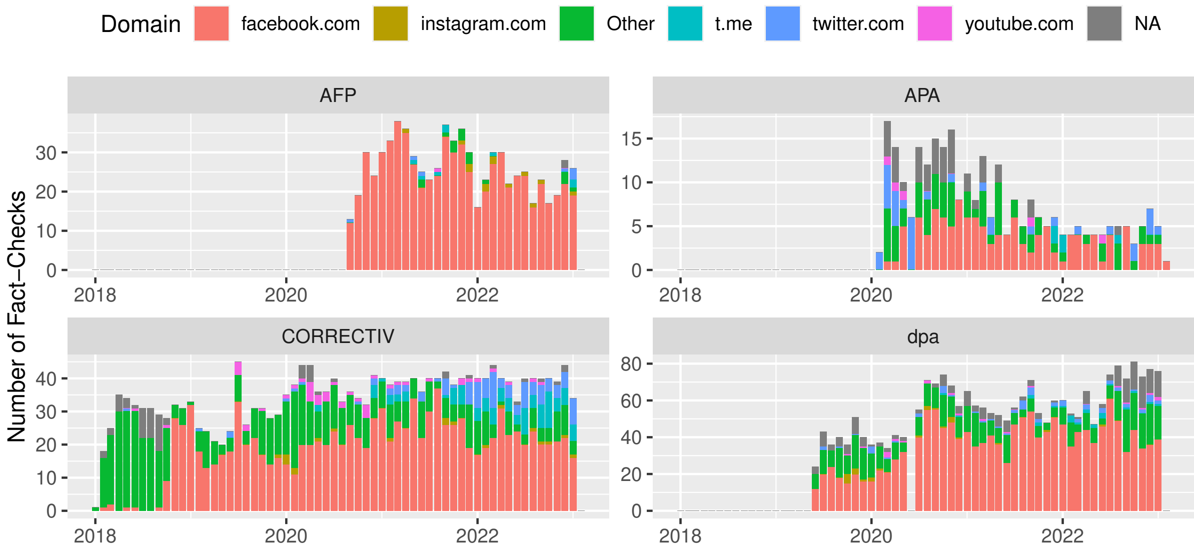

Figure 7: Number of fact-checks per month, organization and the source of the claim.

Domain	AFP	APA	CORR.	dpa
facebook.com	714	130	1156	1579
twitter.com	7	33	108	63
t.me (Telegram)	8	5	95	25
youtube.com	1	5	46	20
instagram.com	13	0	19	24
anonymousnews.org	0	0	27	18
journalistenwatch.com	0	0	21	13
wochenblick.at	1	2	22	4
report24.news	3	4	13	4
reitschuster.de	0	0	9	12
truth24.net	0	0	17	4
Other	11	23	448	226
NA	2	38	53	203
Total	760	240	2034	2195

Table 3: Number of fact-checks per organization depending on the source of the claim.

try often indicates that a fact-check is dealing with a general phenomenon or a claim that is widely spread in different variations. In some cases, it also indicates that the claim was not made by a website or social media platform, for example when politicians make a claim in a public speech.

Figure 7 shows the distribution of claim sources over time for each fact-checking organization. A striking aspect is the almost absolute dominance of Facebook as a source of claims checked by AFP. This contrasts in particular with the APA, which has a greater variance in sources but also does not work with Facebook. They also have relatively more fact-checks with an *NA* entry as the source. The share of Facebook as a source for claims checked by CORRECTIV starts to rise significantly a few months before they start cooperating with Facebook. Nevertheless, both CORRECTIV and dpa also look for other sources of disinformation besides Facebook. Still, the effect of Meta's funding is visible and raises media economics questions about the funding of fact-checking and the incentives that come along.

We also examined which claim sources are associated with particular topics. Figure 8 shows that Telegram has the largest share of the topic *Russo-Ukrainian War*. This supports the findings of a report by the Ukrainian analytical platform Vox Ukraine and its fact-checking section Vox checks, in which the authors show how widespread Russian propaganda is on Telegram (Vox Check, 2022). The other platforms have different focuses: While Facebook, Instagram and Twitter have similar topic shares, the topic *Corona* has by far the largest share on Youtube. The focus on Corona can also be seen on the non-platform domains *report24.news* and *reitschuster.de*, which also have high shares of assignments to the topic *Vaccination*. *Truth24.net* focuses on the topic of *Crimes*, which contains many statements with a xenophobic or racist tone, as it deals with real or faked crimes that are (sometimes erroneously) blamed on migrants.

Reitschuster.de and *truth24.net* also stand out when looking at the ratings given to them by CORRECTIV (see Figure 9). The "lack of context" rating was given relatively more often to the non-platforms than to the platforms whose claims were

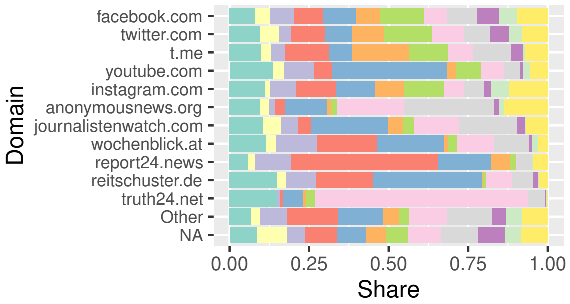

Figure 8: Distribution of the topics in fact-checks depending on the source of the claim; cf., Fig. 2 for legend.

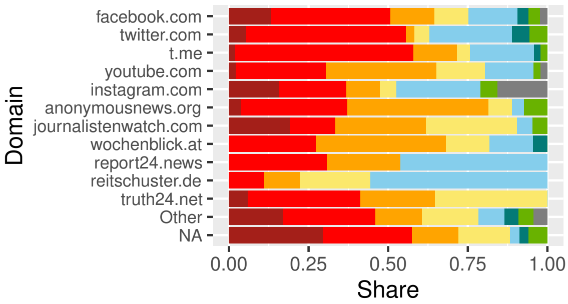

Figure 9: Distribution of the processed merged CORRECTIV ratings (cf., Table 2) depending on the source of the claim; cf., Fig. 6 for legend.

more likely to be rated as incorrect or largely incorrect by the fact-checkers. However, the analysis of claims that do not originate from Facebook should be treated with caution. There are two reasons for this: First, as mentioned above, the number of claims from platforms other than Facebook is much lower, and even lower for the non-platforms. Their observations are therefore much more likely to be highly sensitive to outliers. Second, claims associated with the platforms may have originally been made by other sites that either posted their articles themselves, e.g., on Facebook, or had their articles shared by other users.

4 Conclusion

The topic model analysis using LDA on a dataset of 5229 German-language fact-checks from AFP, APA, dpa and CORRECTIV in the period from 2018 to January 2023 shows that in 2020 all four organizations — unsurprisingly — have a strong focus on (various) Covid related topics. In addition, there is a smooth transition to more mentions of words related to vaccination, resulting in *Vaccination* being the top topic in 2021. Then, at the beginning of 2022, a sudden shift of attention to the Russo-Ukrainian war can be identified. In particular, AFP increasingly combines fact-checks on this topic with visual content checks. At the same time, AFP fact-checks consistently result in negative ratings, and CORRECTIV rarely publishes fact-checks with (partially) positive ratings as well. For the analysis of CORRECTIV's ratings, it is important to merge the ratings of the old and new scales in a meaningful way to avoid false conclusions.

4.1 Discussion

Faceboook claims are clearly checked most frequently (> 68%). The distribution over time suggests that this might also be due to funding from Meta's (now also including Instagram) fact-checking program. Survey data collected from 93 organizations worldwide show that Meta's third party fact-checking program is still the leading funding source in 2022 with 45.2%, while grants cover 29.0% (IFCN, 2023).

This raises several questions: What is the direction of the cause-effect relationships? Is there an unfavourable bias towards current news topics or particular sources? And what consequences can result from this? On the one hand, one can propose that more (independent) money is necessary to ensure a broader attention of the fact-checkers and to slightly loosen the focus from Facebook. It could be a strategic decision that claims that *also* circulate on Facebook are preferably associated with itself. On the other hand, it can be assumed that most claims are in fact circulating on Facebook, so maybe this is not a even a restriction of the thematic range for the general debunking.

4.2 Limitations

The distribution of ratings of AFP shows that often claims are checked for which it is likely in advance that they are false due to the focus on manipulated pictures and videos. This indicates a prioritization of resources and raises the question whether additional financial resources would lead to a better coverage of all *checkworthy* claims, and not only certain misinformation.

In principle, checked sources are still often used as a proxy for topical disinformation spread. Humprecht (2019), for example, uses fact-checks to distinguish between the spread of disinformation in the United States, the United Kingdom, Germany, and Austria. This raises the question to what extent fact-check corpora are representative for disinformation spread. At the same time, there are other approaches to form disinformation corpora, e.g., based on less *trustworthy* sources, identified using NewsGuard[10] scores (Carrella et al., 2023).

Since we focused on topic modeling in the present analysis, the findings are mainly limited to their inductive character (Chen et al., 2023). Nevertheless, we can extract research questions for further analysis.

4.3 Further Research

Further analyses should take into account the challenges and pitfalls of misinformation research (Altay et al., 2023), according to which, for example, misinformation is by no means just a social media phenomenon. Rather, other digital as well as offline media are also prone to misinformation. This is especially important when creating a reference disinformation dataset, which can be used to analyze under-fact-checked topics. By including a reference *quality* media dataset, the relation and the dissemination of (dis)information between low and high quality media can be analyzed. With the help of modern large language models (cf., Grootendorst, 2022; Conneau et al., 2020), it is possible to measure and compare differences in terms of the stance, sentiment and intensity of statements in typical quality media, alternative media, and fact-checks.

Acknowledgements

This research is co-funded by the European Union. We thank the German Press Agency (dpa), the international news agency Agence France-Presse (AFP), the Austrian Press Agency (APA) and the non-profit independent newsroom CORRECTIV for their cooperation regarding the plausibility of the data.

Co-funded by the European Union

References

Sacha Altay, Manon Berriche, and Alberto Acerbi. 2023. Misinformation on misinformation: Conceptual and methodological challenges. *Social Media + Society*, 9(1).

Michelle A. Amazeen. 2020. Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism*, 21(1):95–111.

Marco Bastos. 2022. Editorial: Five challenges in detection and mitigation of disinformation on social media. *Online Information Review*, 46(3):413–421.

David M. Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Roger Blum. 2020. Fakten, Fake News und Wahrheitssuche: Wer checkt die Faktenchecker? pages 251–260. Nomos Verlagsgesellschaft mbH & Co. KG.

Fabio Carrella, Alessandro Miani, and Stephan Lewandowsky. 2023. IRMA: the 335-million-word italian coRpus for studying misinformAtion. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Dubrovnik, Croatia. Association for Computational Linguistics.

Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *NIPS: Advances in Neural Information Processing Systems*, volume 22, pages 288–296. Curran Associates Inc.

Yingying Chen, Zhao Peng, Sei-Hill Kim, and Chang Won Choi. 2023. What we can do and cannot do with topic modeling: A systematic review. *Communication Methods and Measures*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Ullrich K. H. Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga, and Michelle A. Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29.

Fabio Giglietto, Laura Iannelli, Augusto Valeriani, and Luca Rossi. 2019. 'fake news' is the invention of a liar: How false information circulates within the hybrid news system. *Current Sociology*, 67(4):625–642.

Lucas Graves and Federica Cherubini. 2016. *The Rise of Fact-Checking Sites in Europe*. Reuters Institute for the Study of Journalism.

[10]https://www.newsguardtech.com/

Thomas L. Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*.

Dominik Hirndorf and Jochen Roose. 2023. *Welchen Nachrichten kann man noch trauen? Angst vor Desinformation und Vertrauen in öffentlich-rechtliche Medien — repräsentative Umfragen*. Monitor — Wahl- und Sozialforschung. Konrad-Adenauer-Stiftung.

Alexander Hoyle, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Lee Boyd-Graber, and Philip Resnik. 2021. Is automated topic model evaluation broken? The incoherence of coherence. In *NeurIPS: Advances in Neural Information Processing Systems*.

Edda Humprecht. 2019. Where 'fake news' flourishes: a comparison across four western democracies. *Information, Communication & Society*, 22(13):1973–1988.

Edda Humprecht. 2020. How do they debunk "fake news"? A cross-national comparison of transparency in fact checks. *Digital Journalism*, 8(3):310–327.

IFCN. 2023. *State of the Fact-Checkers 2022*. International Fact-Checking Network (IFCN) at Poynter.

Hyun Suk Kim, Yoo Ji Suh, Eun-mee Kim, Eunryung Chong, Hwajung Hong, Boyoung Song, Yena Ko, and Ji Soo Choi. 2022. Fact-checking and audience engagement: A study of content analysis and audience behavioral data of fact-checking coverage from news media. *Digital Journalism*, 10(5):781–800.

S. Lewandowsky, L. Smillie, D. Garcia, R. Hertwig, J. Weatherall, S. Egidy, R.E. Robertson, C. O'connor, A. Kozyreva, P. Lorenz-Spreen, Y. Blaschke, and M. Leiser. 2020. *Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Publications Office of the European Union.

Jianing Li, Jordan M. Foley, Omar Dumdum, and Michael W. Wagner. 2022. The power of a genre: Political news presented as fact-checking increases accurate belief updating and hostile media perceptions. *Mass Communication and Society*, 25(2):282–307.

Philipp Müller and Rainer Freudenthaler. 2022. Right-wing, populist, controlled by foreign powers? Topic diversification and partisanship in the content structures of German-language alternative media. *Digital Journalism*, 10(8):1363–1386.

Nic Newman, Richard Fletcher, Craig T. Robertson, Kirsten Eddy, and Rasmus Kleis Nielsen. 2022. *Reuters Institute Digital News Report 2022*. Reuters Institute for the Study of Journalism.

Katharine Ognyanova, David Lazer, Ronald E. Robertson, and Christo Wilson. 2020. Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School (HKS) Misinformation Review*.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Jonas Rieger. 2020. ldaPrototype: A method in R to get a prototype of multiple latent Dirichlet allocations. *Journal of Open Source Software*, 5(51):2181.

Jonas Rieger, Carsten Jentsch, and Jörg Rahnenführer. 2022. LDAPrototype: A model selection algorithm to improve reliability of latent Dirichlet allocation. *Preprint available at Research Square*.

Lisa Schwaiger. 2022. *Gegen die Öffentlichkeit — Alternative Nachrichtenmedien im deutschsprachigen Raum*. transcript.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS: Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Gerret von Nordheim, Jonas Rieger, and Katharina Kleinen von Königslöw. 2021. From the fringes to the core — an analysis of right-wing populists' linking practices in seven EU parliaments and Switzerland. *Digital Journalism*.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Vox Check. 2022. Disinformation about Ukraine in Russian and pro-Russian Telegram channels.

Nathan Walter, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3):350–375.

Hadley Wickham. 2021. *rvest: Easily Harvest (Scrape) Web Pages*. R package version 1.0.2.

Hadley Wickham. 2022. *httr: Tools for Working with URLs and HTTP*. R package version 1.4.4.