

Best practices for source-based research on misinformation and news trustworthiness using NewsGuard

JULA LUEHRING

University of Vienna & Complexity Science Hub, Austria

HANNAH METZLER

Medical University of Vienna, Austria, Complexity Science Hub, Austria &
Institute for Globally Distributed Open Research and Education

RUGGERO MARINO LAZZARONI

RWTH Aachen, Germany

APEKSHA SHETTY

Complexity Science Hub, Austria & University of Vienna

JANA LASSER

University of Graz & Complexity Science Hub, Austria

Researchers need reliable and valid tools to identify cases of untrustworthy information when studying the spread of misinformation on digital platforms. A common approach is to assess the trustworthiness of sources rather than individual pieces of content. One of the most widely used and comprehensive databases for source trustworthiness ratings is provided by NewsGuard. Since creating the database in 2019, NewsGuard has continually added new sources and reassessed existing ones. While NewsGuard initially focused only on the US, the database has expanded to include sources from other countries. In addition to trustworthiness ratings, the NewsGuard database contains various contextual assessments of the sources, which are less often used in contemporary research on misinformation. In this work, we provide an

Corresponding author: jula.luehring@univie.ac.at

Date submitted: 2024-09-19

analysis of the content of the NewsGuard database, focusing on the temporal stability and completeness of its ratings across countries, as well as the usefulness of information on political orientation and topics for misinformation studies. We find that trustworthiness ratings and source coverage have remained relatively stable since 2022, particularly for the US, France, Italy, Germany, and Canada, with US-based sources consistently scoring lower than those from other countries. Additional information on the political orientation and topics covered by sources is comprehensive and provides valuable assets for characterizing sources beyond trustworthiness. By evaluating the database over time and across countries, we identify potential pitfalls that compromise the validity of using NewsGuard as a tool for quantifying untrustworthy information, particularly if dichotomous “trustworthy”/“untrustworthy” labels are used. Lastly, we provide recommendations for digital media research on how to avoid these pitfalls and discuss appropriate use cases for the NewsGuard database and source-level approaches in general.

Keywords: misinformation, source trustworthiness, NewsGuard

Misinformation spreads on both social media platforms (Lazer et al., 2018) and mainstream media channels (Tsfati et al., 2020). However, only a fraction of news consumption is attributed to false news – roughly 0.15% (Allen et al., 2020) – and only a minority of users engage with unreliable information online (Baribi-Bartov et al., 2024; Grinberg et al., 2019). Typically, these individuals hold strong political or ideological beliefs and accept and actively look for information that echoes their opinions (Ecker et al., 2022) – information that often comes from political elites or established media sources (Tsfati et al., 2020). Misinformation capitalizes on these dynamics to stoke conflict and evoke negative emotions, thereby exacerbating partisan divides and fueling inter-group hostility (González-Bailón et al., 2023; Robertson et al., 2023). Against a backdrop of declining trust in both media outlets (Newman et al., 2023) and institutions (Bennett and Livingston, 2020), it becomes increasingly crucial to comprehend the mechanisms driving the dissemination of misinformation, particularly in the digital sphere. While digital media facilitates the measurement of these dynamics, it also has the potential to reinforce polarization and misinformation (Lorenz-Spreen

et al., 2022).

Scholars across various disciplines have extensively studied the phenomenon of misinformation, employing terms such as misinformation, disinformation, fake news, rumors, or conspiracy theories, often used interchangeably (Guess and Lyons, 2020). Although these terms share a focus on veracity, they differ in terms of intentionality. Misinformation is commonly used as an umbrella term for inaccurate or unreliable information, regardless of intent, e.g., Ecker et al. (2022) or van der Linden (2022). However, the ambiguous nature of misinformation presents significant methodological challenges, granting researchers several degrees of freedom. Defining the boundaries of misinformation — essentially determining where the truth ends — is a complex task for several reasons. First, misinformation often appears in subtle forms rather than obvious lies (Allen et al., 2024; Altay et al., 2023; Bakir and McStay, 2018; Vargo et al., 2018), making it hard to define where it starts. Implicit, often partisan, interpretations of facts differ from explicit lies, making subtle forms of misinformation hard to detect. Second, identifying misinformation involves ethical considerations, with different experts reaching different conclusions. To navigate these challenges, researchers often focus on binary (true/false) cases of misinformation, typically identified by fact-checkers. While this approach addresses some ethical concerns by clearly classifying false information, it captures only a small portion of misinformation. Finally, conceptual choices are often shaped by methodological constraints (e.g., available data) or disciplinary preferences (e.g., methodological traditions), which is why misinformation is often used as a general umbrella term for different phenomena (Weeks and Gil de Zúñiga, 2021). Consequently, this semantic and methodological variability has contributed to imprecise or inconclusive findings regarding the prevalence and impact of misinformation. While some scholars, such as Altay et al. (2023) or Budak et al. (2024), argue that the effects of misinformation are overstated, calling it a symptom of broader social issues, others, like Tay et al. (2024), reason that the heterogeneity in measuring misinformation contributes to divergent evidence and creates a false dichotomy between understanding misinformation as a symptom or a cause (also see Bozarth et al., 2020).

In digital media research, one of the most prominent ways to measure misinformation has been to use a list of sources that are known to share misinformation frequently. This approach has the advantage of being scalable, however, it is important to note that it addresses the editorial practices of the sources rather than the veracity of the content itself. NewsGuard has emerged as the most widely used database for source ratings (Aslett et al., 2022; Guess et al., 2020, 2021; Pratelli

et al., 2023; Robertson et al., 2023) and is particularly popular in misinformation research for its comprehensive source-level trustworthiness scores, which are compiled based on web tracking data and rated and updated by professional editors and journalists. In addition to its comprehensiveness, NewsGuard also offers a more fine-grained assessment of the trustworthiness of sources based on adherence to journalistic quality criteria and provides a point score ranging from 0 (the source is untrustworthy because it severely violates basic journalistic standards) to 100 (the source adheres to all standards of credibility and transparency). Since the start of the database in 2019, NewsGuard has constantly added new sources. While initially the focus of NewsGuard was the US, the database has expanded to other countries, such as Germany and France. Given the popularity of the NewsGuard database as a tool to quantify the prevalence of misinformation, an in-depth analysis of the temporal stability of ratings and the completeness of the database for different countries is warranted. Furthermore, NewsGuard requires a license, and researchers in need of a measurement instrument to quantify misinformation must weigh the advantages and disadvantages of source-based methods and the investment required to access the NewsGuard database. With the present in-depth analysis of the database's content, we aim to provide researchers with the necessary information to make this decision. Where warranted, we provide recommendations for effectively utilizing the NewsGuard database and source-based approaches alongside our analysis of the content of the database.

Tracking misinformation on digital media

To study the diffusion or discussion of misinformation on digital platforms, computational researchers commonly opt for either content- or source-based methodologies, searching for web links to content (via the URL) or a source (via their web domain). Both approaches rely on experts, such as fact-checkers, to curate lists of content or sources, which serve as the foundation for collecting data from digital platforms. Source-based strategies, which are the focus of the present work, offer the advantage of encompassing a broader spectrum of sources and narratives, thus enriching the sample by considering a more diverse information environment. Source-based approaches also enable the inclusion of subtler forms of misinformation and facilitate accurate estimations of both the scale (Allen et al., 2020; Grinberg et al., 2019; Guess et al., 2020; Yang et al., 2021) and dynamics of dissemination (Lasser et al., 2022, 2023; Guess et al., 2021; Robertson et al., 2023; Shao et al., 2018).

The precision of source-based approaches hinges mainly on the selection and quality of

source ratings within a given list. Strictly speaking, this conceptualization covers information disseminated by unreliable sources (rather than individual pieces of misinformation). Therefore, observed patterns may be intricately linked to the inherent biases or external forces in the selection of sources. For instance, if we analyzed misinformation in politicians' social media posts, an unbalanced selection of unreliable sources would most likely reveal patterns due to partisanship rather than patterns universal to misinformation. This underscores the importance of assessing confounding factors such as political biases or editorial practices of the sources, considering that a lack of editorial oversight can compromise the quality of news reporting (Lazer et al., 2018). In other words, to obtain a balanced selection of sources and account for the influence of source-related variables other than reliability, source-based approaches can incorporate assessments of political, cultural, or editorial biases and misleading tendencies while maintaining generalizability. Nonetheless, several problems remain in practice: First, source categorizations often rely on a binary classification system (fake/real), oversimplifying the nuanced spectrum of misinformation. Second, it is crucial to acknowledge that source ratings vary across time and cultural contexts, for instance, following a change in ownership or editorial practices. For example, Lin et al. (2024) combined source-level quality ratings with context-specific keywords, e.g., hashtags linked to an anti-vaccination protest in Ottawa to account for misinformation-sharing in different populations. Third, while source judgments are typically made without requiring specific topic expertise, they often reflect the journalistic traditions of a particular media ecosystem, which may not be universally applicable across different contexts (Brüggemann et al., 2014). Source-based approaches are a promising way to track online misinformation at scale. However, to study the spread of untrustworthy news, scientists need access to reliable and credible databases of source ratings.

NewsGuard: A popular database for source ratings

Since the 2016 US elections, the most popular way to track misinformation online has been to use a list of untrustworthy or unreliable sources and search for their web domains in URL text (Lasser et al., 2022; Grinberg et al., 2019; Guess et al., 2020, 2021; Shao et al., 2018; Yang et al., 2021). The organization NewsGuard offers the most comprehensive list of such domains (Lin et al., 2023). It does not just provide a blacklist of untrustworthy sources but provides a trustworthiness score for each source, covering news sources across the entire spectrum of news quality. A trained team of experts, mainly consisting of journalists and editors, rates the news sources based on

nine journalistic quality criteria¹ and regularly updates both the coverage of the list by adding new sources and the trustworthiness ratings of existing sources. Since its start in 2019, the database has been constantly growing and has recently extended to countries beyond the United States, such as Germany and Italy. With recent political events and crises, demand for such databases has increased, enhancing the popularity of NewsGuard. For instance, it has been used to study user engagement with unreliable news during the 2016 and 2020 US elections (Pratelli et al., 2023; Robertson et al., 2023), to assess the changing quality of cited information of political elites in the US (Lasser et al., 2022) or alternative understandings of honesty and truth in political discourse (Lasser et al., 2023), and to estimate the effects of source labeling on decreasing misinformation sharing (Aslett et al., 2022; Celadin et al., 2023).

However, the NewsGuard database was initially created as a tool for brand security rather than scientific research. In addition, employing journalists and editors to assess the sources included in the list incurs significant expenses. Given its predominantly private audience and expenses, the NewsGuard database is only available to subscribers who pay to license the database. Here, we provide a comprehensive assessment of the database to help misinformation researchers weigh the resources required to access the NewsGuard database and the advantages and limitations of source-based approaches in general. We aim to answer the following main research question: How volatile are source-level trustworthiness ratings, such as those in the NewsGuard database, across time and language contexts? To answer our research question, we first assess the distribution of trustworthiness ratings, the process of updating them, and the inclusion and removal of sources over time. Second, we assess the completeness of the database for different countries. Lastly, we estimate the usefulness of other source-level information provided in the database, such as the political orientation and topics covered by sources. To assess different aspects of the database, we introduce three sub-questions:

How stable are trustworthiness ratings over time? Investigating the development of the NewsGuard database over time is important to estimate whether using an older version of the database distorts results and should be used to measure trustworthiness in older data. We describe the composition of the trustworthiness score and different database versions. We also explore the frequency and reasons for major changes, such as how often ratings change or sources are removed.

¹<https://www.newsguardtech.com/ratings/rating-process-criteria/>

Additionally, we provide an exemplary reproduction of published research using NewsGuard ratings from different points in time to assess the degree to which results vary based on changes in the measurement instrument.

How complete is the database across countries, particularly for non-US countries? A recent study has shown high agreement of NewsGuard with other expert-rated lists in the US context (Lin et al., 2023), but an assessment for other countries is missing. Therefore, we describe the coverage across and differences between countries and assess the stability of source ratings and coverage over time. Further, we compare the completeness of information provided by NewsGuard across different contexts by cross-checking the NewsGuard database with a dataset of domains shared on social media for the US, Germany, and UK, and other existing lists of low-quality web domains.

How valuable are contextual source labels for misinformation research? NewsGuard labels the topics covered as well as the political orientation sources. We analyze the temporal stability and completeness of these labels across countries. Additionally, we manually validate the labels for German-speaking sources. We assess in what way such labels can complement source characterizations and shed light on aspects of misinformation beyond trustworthiness.

Dataset construction

NewsGuard curates the list of sources it rates by tracking online activity, using multiple media monitoring tools as well as their own processes to monitor specific topics and social media platforms. Per country, NewsGuard employs a group of editors to review the sources, including a rapid response team. The organization also manually adds sources they deem influential or that changed their web domain. On their website, NewsGuard claims to have reviewed news sources accounting for 95% of online engagement, including news consumed and shared online. To avoid bias and inconsistency, NewsGuard applies multiple layers of editorial review. The database automatically updates every hour and allows tracking changes back to March 2019, when the organization started operating. Still, how domains are selected or removed is not disclosed on the website. As of September 15, 2024, the database has 12,288 entries in total, some without a trustworthiness rating (7.6%). This includes 70 sources classified as a “platform” (e.g., YouTube), 63 sources judged as “satire” (e.g., The Onion) and 806 lifestyle sources (e.g., healthquote-free.com). For the remaining

11,349 sources, the database includes the following characteristics:

- Unique identifier for each rating and date of the last update
- Domain name and parent domain (e.g., nytimes.com)
- A trustworthiness rating as a quasi-continuous scale (between 0 and 100) and binary labels (N = Not Trustworthy/T = Trustworthy) based on a threshold at 60
- Language and country associated with the source
- Political orientation of the source (e.g., “left” and “right”)
- Topics covered (e.g., “local news”, “political news”, “health information”) by the source

The team of NewsGuard experts regularly judges the trustworthiness of the selected news sources based on nine journalistic criteria that are universally applied across countries. The trustworthiness score is a composite score of these nine indicators. More precisely, experts assign a binary label for each criterion. Each criterion has a weight, a number of points, which together make up the overall trustworthiness rating (ranging from 0 to 100). Here, we list the criteria and their weights in descending order²:

1. “Does not repeatedly publish false content”, changed to “Does not repeatedly publish false or egregiously misleading content” in December 2023 (22 points)
2. “Gathers and presents information responsibly” (18 points)
3. “Regularly corrects or clarifies errors”, changed to “Has effective practices for correcting errors” in December 2023 (12.5 points)
4. “Handles the difference between news and opinion responsibly” (12.5 points)
5. “Avoids deceptive headlines” (10 points)
6. “Website discloses ownership and financing” (7.5 points)

²<https://www.newsguardtech.com/ratings/rating-process-criteria/>

7. “Clearly labels advertising” (7.5 points)
8. “Reveals who’s in charge, including possible conflicts of interest” (5 points)
9. “The site provides the names of content creators, along with either contact or biographical information” (5 points)

Since the overall score is a sum of the individual scores for each criterion, it is not truly continuous, as some values (for example, 1, 2, 3, or 4 points) cannot be achieved by any combination of criteria.

How stable are trustworthiness ratings over time?

In the following, we review how the database, particularly the trustworthiness ratings and source selection, have changed over the years since the first version of the database was compiled in 2019.

Coverage across time. Over the span of five years, NewsGuard has grown from 2,375 to 12,288 entries in total, as shown in Figure 1A to C. NewsGuard regularly adds new sources but rarely removes any, i.e., the database has been growing to four times its initial size. When counting the added domains based on monthly comparisons of the list, NewsGuard has added 8,906 domains over time but only removed 685, adding an average of 137 new domains per month.

NewsGuard ratings are available for customers in an online database (Amazon S3 bucket), where a current snapshot is uploaded every hour. We sampled a snapshot from the first hour of the 15th day of every month. When we instead sampled from the second day of each month, we observed two irregular dips in the number of sources (see May 2022 and February 2024 in Figure 1A). It is likely that errors caused this fluctuation. For instance, in May 2022, the approximately 3,500 removed sources that did not appear in the database were present in the database two weeks later. Yet, this sudden disappearance and reappearance of sources illustrate the overall volatility of the database. We therefore advise researchers to inspect several database snapshots and check their consistency before choosing one.

The most recent snapshot contains ratings for 11,349 unique domains in the 12,288 total entries; indicating that some sources appear multiple times. Typically, these sources publish in dif-

ferent languages, i.e., they have the same web domain but provide multi-lingual content, or they have an entry as a local and a global source. For instance, the Austrian website servustv.com produces both German- and English-language output and appears with the country label “AT” and the label “ALL”. According to NewsGuard, the latter serves as a generic label for the English-language translation of sites that were also rated in their respective language. In other words, each site with a label “ALL” also exists in at least one other language, and the label “ALL” can largely be ignored. In line with that, duplicated sites have identical trustworthiness ratings across country or language categories, except for three cases. Therefore, dropping duplicated domain entries does not substantially affect downstream research results. However, if a classification of sources by country is necessary for the research question at hand, researchers should keep duplicated domains.

Furthermore, 5,754 of the rated domains belong to only 444 unique parent domains. Some parent domains have multiple variations of the domain name, probably representing the same source, e.g., theragingpatriot.org has domains ending with .blog, .com, .net, and .pro. Other parent domains have very different sub-domains, e.g., 24usnews.com and africdaily.com belong to charmdaily.com. The other 6,534 ratings do not have a parent domain entry in the database. Domains belonging to the same parent domain usually have the same trustworthiness rating. Furthermore, NewsGuard updates domains belonging to the same parent domain at the same time, and they seem to assign the same unique identifier to those domains. Therefore, while the database contains 10,862 unique domains, it only contains 7,073 unique identifiers and, therefore, individually rated (parent) domains.

Distribution. NewsGuard trustworthiness ratings range from 0 to 100. Historically, NewsGuard considered sources with a rating below 60 as not trustworthy, therefore assigning binary trustworthiness labels³. Since February 2023⁴, NewsGuard uses a more granular categorization, classifying sources with a score of 100 as “high credibility”, scores between 75 and 99 as “generally credible”, scores between 60 and 74 as “credible with exceptions”, scores between 40 and 59 as “proceed with caution”, and scores between 0 and 39 as “proceed with maximum caution”. Notably, NewsGuard is not a blacklist. In fact, close to 60% of sources have an overall rating of over 60 (see Fig. 1B for the distribution of trustworthiness scores in the most recent vs. the first version of the

³Snapshot of the described rating process from December 31st, 2022: <https://www.newsguardtech.com/ratings/rating-process-criteria/>

⁴Snapshot of the updated rating process from February 1st, 2023: <https://www.newsguardtech.com/ratings/rating-process-criteria/>

database and Fig. A.1 in the appendix for distributions per year). Furthermore, the distributions also clearly show a multimodal structure, indicative of the quasi-continuous nature of the overall News-Guard score. In addition, some combinations of characteristics seem much more prevalent than others, indicating that the individual characteristics are not independent (see also “Composition of trustworthiness score” below).

When looking at the score distribution over time, the skew towards high ratings diminishes slightly: while the average score was 71.8 ($SD=33.3$) in 2019, it is now at 63.6 ($SD=32.6$) as shown in Figure 1C. Various explanations for decreases in average trustworthiness over time are possible: First, the database has quadrupled in size and added many sources with lower trustworthiness scores (on average 59.6, $SD=27.5$).

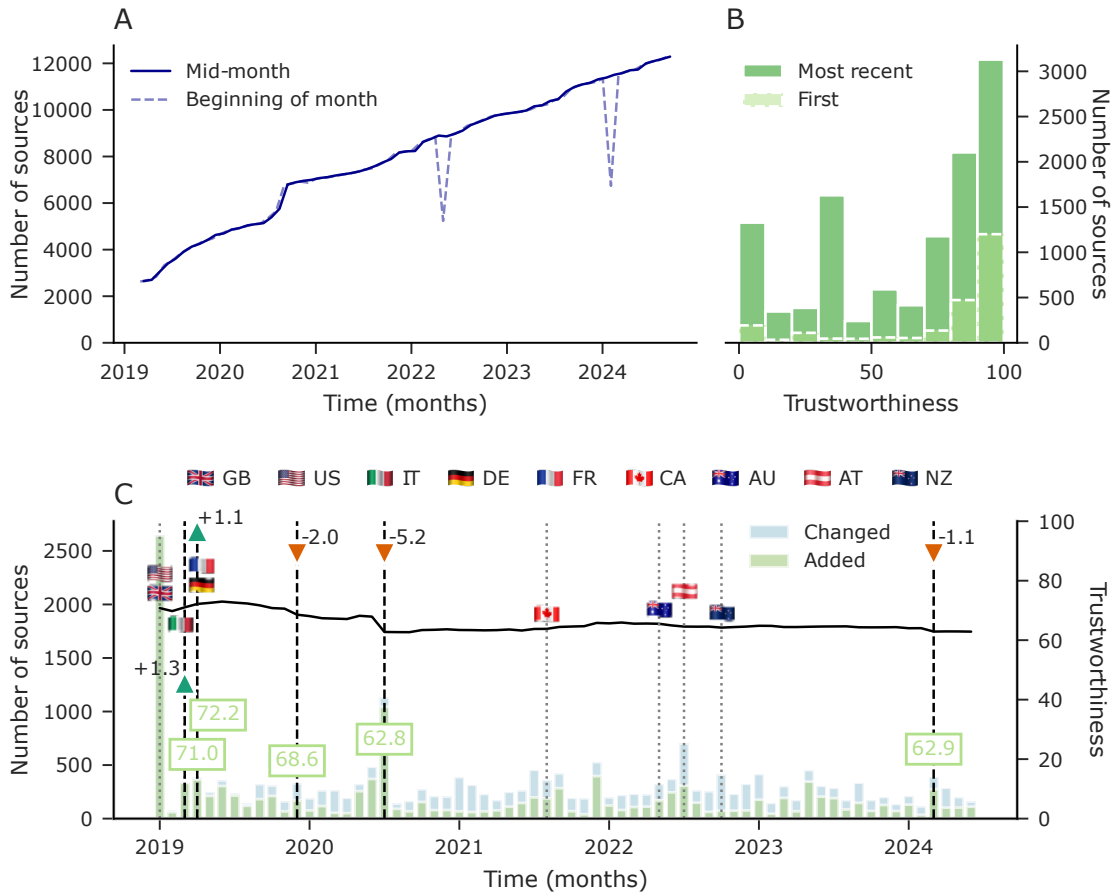


Figure 1. Description of trustworthiness ratings with panel A showing the number of sources with rating over time based on two different sampling strategies: when sampling one snapshot of the database mid-month (solid line), we observe a steady increase, whereas sampling at the beginning of the month results in two irregular dips (dashed line). Panel B: Distribution of trustworthiness in the first and most recent database as a histogram. Panel C: Changes in trustworthiness scores over time (monthly granularity). Flags (respective country codes in order of appearance) and dotted lines indicate when countries were added and dashed lines with green and red arrows highlight the top five major score changes. The height of the bars describes the number of sources added vs. changed, with colors indicating the proportions. We also show the average trustworthiness of sources added in the green text boxes.

A significant number of sources with lower ratings were added, in particular during the onset

of the COVID-19 pandemic in 2020. The dashed lines with arrows in Figure 1C show the top five updates in terms of their impact on the average trustworthiness. The height of the bars shows the number of sources added vs. changed (with proportions in colors and the average trustworthiness of added sources in green). For all major updates, the proportion of added sources is larger than changed sources. The highest number of sources added was in September 2020, when 1,030 sources with on average very low trustworthiness ($M=35.3$, $SD=14.1$) were included. It is unclear if this development is due to an effort by NewsGuard to include more untrustworthy news or a global increase in untrustworthy sources.

Second, NewsGuard employs teams of experts per country, potentially leading to differences in coverage and ratings between countries. Dips in trustworthiness may be related to country-specific sources being added. For instance, 1,015 of the 1,030 added sources in September 2020 are coded as US, resulting in a drop in average trustworthiness of roughly 5%. In contrast, adding entirely new countries to the database seems to have little effect on the overall trustworthiness score (see the dotted lines with country flags in Figure 1C), indicating that NewsGuard curates a balanced mix of trustworthy and untrustworthy sources when adding a new country. Overall, smaller dips appear across all countries covered and are most likely related to many low-quality sources gaining traction around the time of the COVID-19 pandemic.

A third influence on the overall trustworthiness score is updates of existing ratings. NewsGuard indicates the date when a rating was last updated, even if the score did not change. Updates to entries are documented on average once every eight months (249 days), but this varies strongly per source between a few seconds to a maximum of 790 days (interquartile range: 157 to 334 days). When updated, trustworthiness scores mostly remain unchanged (on average, in only 20% of updates the rating is changed). In other words, ratings realistically change on average every two years ($M=570$ days, $IQR=397$ days to 670 days). In addition, the majority of sources have not decreased in trustworthiness (average change of rating per source= 0.1 , $SD=3.5$). However, a few sources drastically dropped in trustworthiness; for instance, *conservativedailynews.com* lost 80 points in one update in November 2020. Figure A.2 in the appendix illustrates the differences in consecutive scores over time, with major score updates in early 2020.

Overall, the decrease in the overall trustworthiness of sources we can see in Figure 1C can, therefore, be attributed to NewsGuard adding untrustworthy sources rather than previously

trustworthy sources losing points. This interpretation is also supported by our analysis of changes in individual rating criteria below.

Composition of trustworthiness score. The trustworthiness score is a composite score calculated based on nine journalistic criteria. For each source, NewsGuard’s raters assign a Boolean value (yes/no) per criterion. The trustworthiness score is then calculated as the sum of the weighted criteria. Further explanation of the criteria and their weights can be found in the NewsGuard FAQ⁵.

In the latest version of the dataset, 81.3% of sources fulfill the criterion “Avoids deceptive headlines” (9,217 sources). The second most prevalent criterion is “Does not repeatedly publish false or egregiously misleading content”, which is met by 81.2% of all sources. These two top criteria often go together: When one is satisfied, the other tends to be as well (Pearson’s correlation coefficient = 0.9). A factor analysis with oblique rotation confirms that the two criteria load on one factor, accounting for 58.1% of the explained variance, with factor loadings > 0.8. All other criteria share 36.9% of the variance explained (factor loadings > 0.5). In line with that, 11.8% of all sources meet only the two top criteria combined, while 13.4% of sources satisfy none of the nine journalistic criteria. The most common combination of criteria is that all nine criteria are satisfied simultaneously (13.5% of sources). We provide a correlation matrix of the criteria in the Appendix (Table A.1).

We find that around 60% of sources have never changed in their fulfillment of journalistic quality criteria during updates. We present update patterns of individual criteria in Figure 2A. The most common change is sources stopping disclosing ownership and financing, accounting for 12.3% of the observed changes. The opposite change (a disclosure of ownership and financing) is the third most common pattern, occurring in 10.7% of score updates. This indicates either that NewsGuard frequently checks criteria related to ownership or that adherence to such criteria is more volatile.

⁵From December 31st, 2023: <https://www.newsguardtech.com/ratings/rating-process-criteria/>

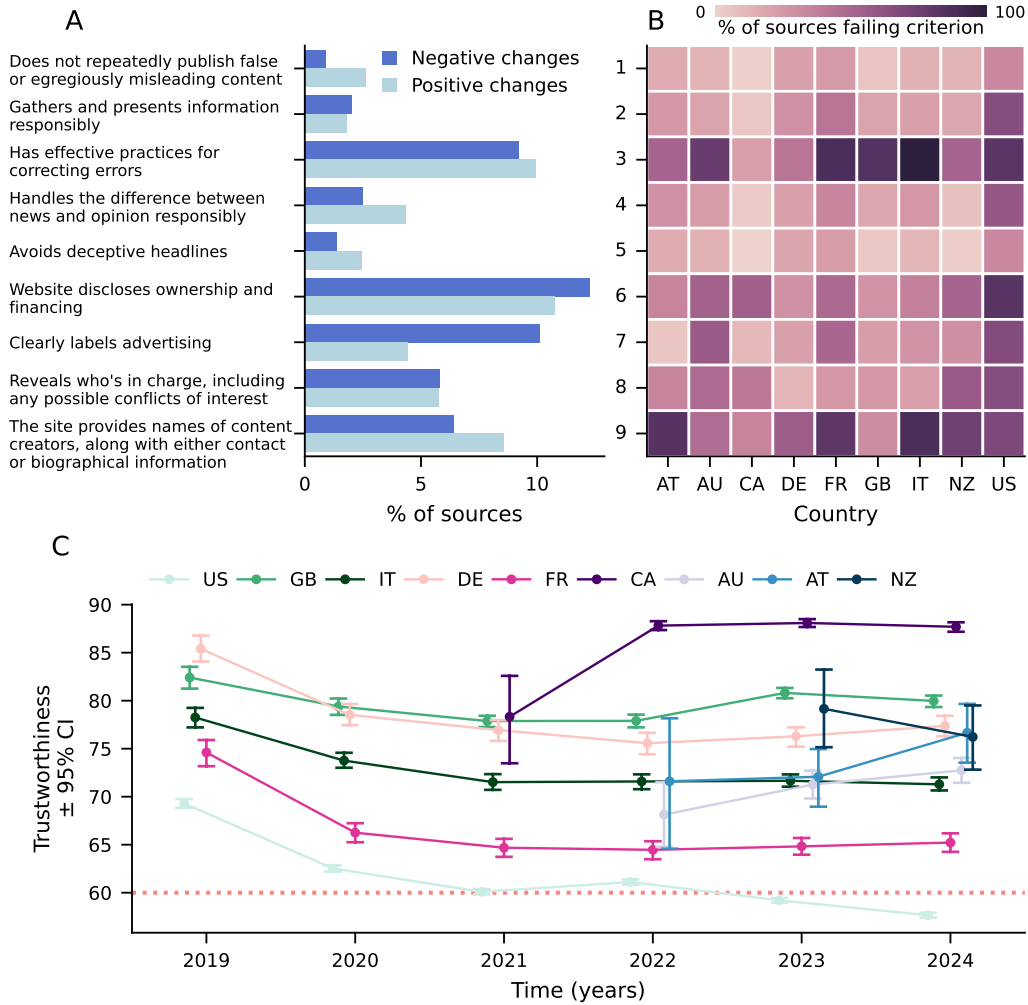


Figure 2. Panel A: Percentage of sources that have stopped or started to fulfill a criterion, including multiple changes (positive vs. negative) of a single source. Panel B: Heatmap showing the percentage of sources that do not fulfill a criterion per country as of September 15th, 2024 (the numbers represent the criteria as shown in panel A). Panel C: Average trustworthiness per country over time (truncated to range from 55 to 90), aggregated per year with 95% confidence intervals.

Reproducibility of downstream research results using different snapshots of the database.

Over the observation period, we observe both a substantial addition of sources, as well as more minor changes in the fulfillment of individual journalistic criteria and therefore, the overall score for individual sources. This begets the question of whether such changes can potentially influence the conclusions drawn from research that rely on the NewsGuard database, depending on which version is used. To provide a partial answer to this question, we reproduce some of the analyses of the trustworthiness ratings of sources shared by politicians in the US, UK and Germany by Lasser et al. (2022), using different versions of the NewsGuard database. In Figure 3A, we show the temporal development of the average NewsGuard score of sources shared by Democrat and Republican members of Congress on Twitter between 2016 and 2022. We use snapshots of the NewsGuard database taken on March 1 from 2019 to 2024 (the original research used the snapshot from March 1, 2022). In addition, we show averaged NewsGuard scores for observations that are dynamically matched with the scores in the NewsGuard database on a monthly basis (dotted lines). For example, the data analyzed for March 2020 is paired with the March 2020 version of the NewsGuard database. For Republicans, the result barely changes between different versions of the database, except in 2019, where scores are substantially higher – on average 3.2 points – than the snapshot from 2020. The differences between snapshots from other years do not surpass 0.3 points and reveal no trend. For Democrats, the snapshot from 2019 also has substantially higher scores (2.1 points when compared to 2020). In addition, the snapshot from 2024 yields a 1.3 points lower average score than the snapshot from 2023. For other years, the differences also do not surpass 0.3 points and have no trend. However, even given the sometimes relevant differences in scores, the main finding of the research, namely the increasing difference in average scores between Democrats and Republicans, does not change.

Figure 3B shows the average difference in the NewsGuard score of sources shared by the politicians over time. When using the 2019 snapshot, which deviates most strongly from the others, the difference between parties is smaller in the years 2020 to 2022 compared to other snapshots, but remains substantial. The finding that there are only minor changes in results depending on which version of the database is used for the US for the continuous scale is likely due to the fact that sources rarely experience large changes in their rating, and that almost all major US news sources were already covered in the first version of the NewsGuard database: the snapshot from 2019 covers 15.0% of links posted by Democrats on Twitter and 18.5% of links posted by Republicans. For the 2024 snapshot, the coverage only increases slightly to 16.3% and 19.6% of links for Democrats and

Republicans, respectively. The use of a dynamic matching of NewsGuard database version does not change the picture as long as continuous scores are used (see more on this below) and coverage has stabilized.

For the other two countries, Germany and the UK, the picture is somewhat different: differences in average ratings stratified by party show larger deviations between years (see Figures A.3 and A.5 in the appendix). This is likely due to massive changes in the coverage of the sources shared by politicians in these countries by NewsGuard. For example, for members of the German party SPD, coverage increased from 0.5% in 2019 to 12.7% in 2024 and from 0.2% to 12.0% for members of CDU/CSU. In the UK, the coverage for members of the Labour Party increased from 7.8% to 13.5%, and for Tories, from 2.4% to 7.8%. Taken together, such coverage differences between years would lead to different conclusions in some cases, but not in others. For instance, the NewsGuard score of small parties like DIE LINKE and the Greens in Germany shows an upward trend with the 2019 version but not with newer versions. In the UK, all versions would yield the same main conclusion, namely that the average score shows no specific trend over time. It is usually the earlier versions of the database that lead to different scores, given that coverage was still low when countries were first included. Later versions, especially 2022-2024, usually yield the same result.

As the difference between years is largely caused by differences in coverage and not by differences in the ratings of individual sites, it is questionable if the use of a dynamic matching of scores and observations would increase the validity of results: Observations from earlier times with lower NewsGuard coverage of sources would be less likely to be matched with a score, introducing a higher likelihood of bias for these times due to uneven coverage. Using more recent snapshots of the NewsGuard database to infer ratings at earlier points in time – as is effectively done if, for example, a snapshot from 2024 is used to score news sources in data collected in 2019 – alleviates the lack of coverage at the potential cost of using ratings that are less accurate. In summary, our analysis shows that for countries other than the US, even for continuous scores, the usage of different versions of the database has the potential to significantly influence research outcomes, likely due to the addition of a significant number of relevant new sources to the database over the years. Researchers investigating countries for which sources were only recently added to the database (e.g., New Zealand or Austria) should, therefore, proceed with caution and invest time to validate the coverage of the database for a given country.

Figure 3C shows the fraction of untrustworthy links, e.g. sources with a NewsGuard score <60 that were shared by US politicians. This is a common approach for creating a binary “trustworthy” vs. “untrustworthy” label from NewsGuard scores. Different from the results shown in panel A, panel C shows that depending on the snapshot of the database that is used, there are large differences in the percentage of shared links that are considered untrustworthy – particularly for Republicans. For example, the earlier NewsGuard versions from 2019 and 2020 indicate a relatively stable fraction of untrustworthy links between 2% and 3%, whereas later versions show a rising fraction that reaches 8% from 2020 onward. Analyzing this further, we find that these changes are largely driven by a few large news sites, mostly popular with conservative politicians, crossing the cutoff threshold of 60 and becoming classified as “untrustworthy”: In 2019, the sites “thefederalist.com” and “dailywire.com” changed to “untrustworthy” – sites that are linked to a total of 771 and 547 times in the observation period, out of which 99.4% and 99.3% of links come from members of the Republican party, respectively. Note that we only report changes for sites that were linked to more than 300 times in the observation period. Similarly, in 2020 “breitbart.com” (2504 links, 99.3% Republican), “washingtontimes.com” (1790 links, 94.0% Republican), and “newsmax.com” (546 links, 97.1% Republican) changed labels. The only change to “untrustworthy” of a major site that is predominantly linked to by Democrats occurs in 2022 for “msnbc.com”, which accounts for 1176 links (94.3% Democrat). Another enormous change would have become apparent in late 2022, as “foxnews.com” changed to “untrustworthy” for four months (August to November). As this time is outside our observation period for this replication study, the change does not become visible in Figure 3C. However, “foxnews.com” alone accounts for 13,124 links (3.1% of all links with a NewsGuard score) in the data set, out of which 97.4% come from Republicans, illustrating further how the rating of a single news site can substantially shift the obtained results if a binary classification is used.

Looking at other countries, the difference in the fraction of untrustworthy links detected depending on the snapshot of the database used becomes even more striking: for Germany, if the 2024 snapshot of the database is used, the fraction of untrustworthy links shared by members of the FDP and CDU increases from a negligible amount to around 5%, while the share of untrustworthy links shared by members of the AFD doubles from around 10% to around 20% (see Figure A.4 in the appendix). This change is entirely driven by one of the largest German newspapers, “bild.de”, crossing the 60 points threshold. Of the links to “bild.de”, 54.9% come from members of the AFD, 25.6% from the CDU and 10.3% from the FDP, with the remaining 9.2% divided among members

of the parties to the left of the political spectrum.

While we do not observe similarly extreme changes for the UK (see Figure A.6 in the appendix), the finding that a binary classification of US and German news sources into “untrustworthy” and “trustworthy” can lead to large differences in the fraction of detected untrustworthy links for different versions of the database is concerning. This includes major differences in the interpretation of the derived results, such as whether there is a trend detected or not, and whether members of a party do share links to untrustworthy sites at all. In summary, we conclude that the use of binary ratings has the potential to substantially distort research findings and continuous ratings should be used wherever possible. In addition, large changes in the prevalence of “untrustworthy” sources should always be analyzed in detail to see if they are a general phenomenon or can be tied to a single major source changing classification. In cases where no continuous ratings are available, the use of dynamic ratings could somewhat alleviate the situation, as the dichotomous dynamic ratings would at least correspond more closely to the journalistic standards at a given point in time.

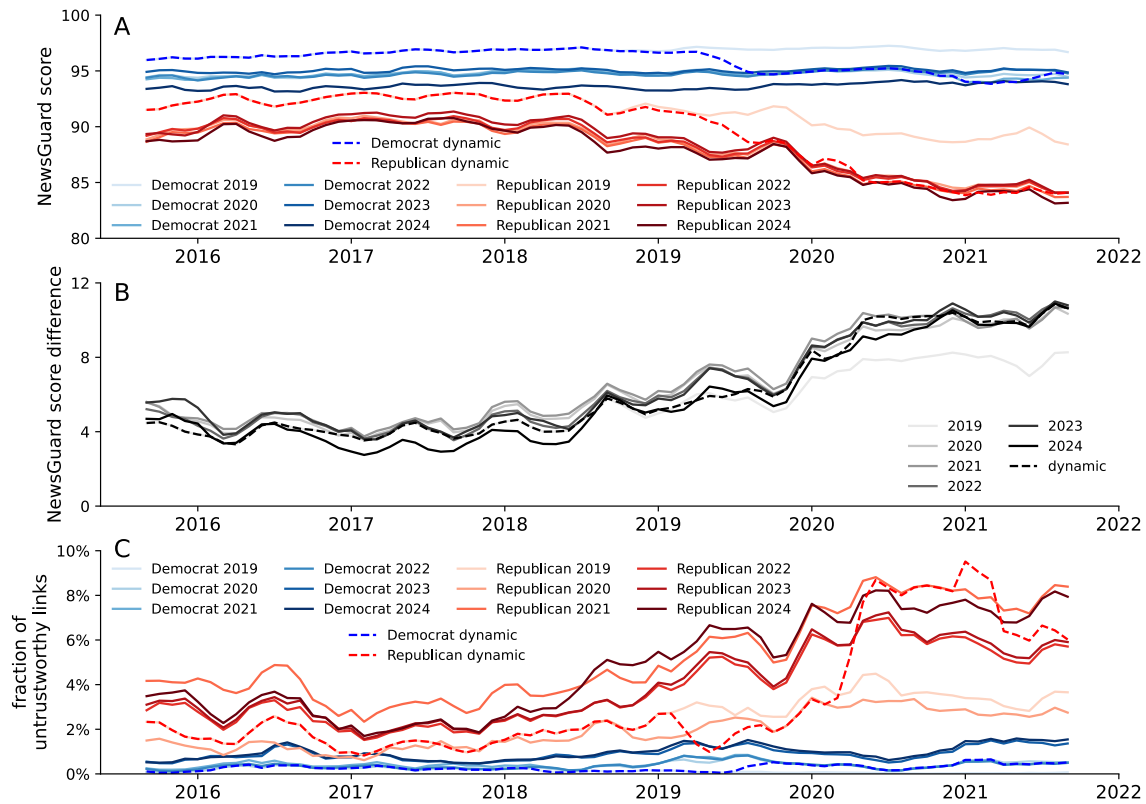


Figure 3. Reproduction of the time series of NewsGuard scores of US Congress Members from Ref. Lasser et al. (2022) using different versions of the NewsGuard database from 2019 to 2024 (always taken from the first hour of March 1st of the given year). Panel A: Average scores for Democrats and Republicans indicated in different shades of blue and red for each year. Panel B: difference between the average Democrat and average Republican score, indicated in different shades of gray for each year. Panel C: Fraction of untrustworthy links (NewsGuard score < 60) with different versions of the NewsGuard database (as in panel A). For each panel, the dotted lines indicate a dynamic value, where data was taken from the version of the NewsGuard database corresponding to the year and month of the observation. Note that as the first snapshot of the NewsGuard database is available for March 2019, all observations before that point in time are matched to the NewsGuard version from March 2019. The time series represent a moving average over three months.

How complete is the database across countries, particularly for non-US countries?

Country comparisons. The dataset (2019-2024) contains sources from nine countries in total (United States, Great Britain, Italy, Canada, France, Germany, Austria, Australia, and New Zealand) plus sources considered to be global. Countries were added in the following order: Great Britain (country name in the database “GB”), United States (US), Italy (IT), France (FR), Germany (DE) – all added in 2019, Canada (CA, added in 2021), Australia (AU, added in 2021), Austria (AT, added in 2022), and New Zealand (NZ, added in 2023). In the following analysis, we will drop the category “ALL”.

The majority of sources are from the United States (76.1%), followed by Great Britain (5.2%). Table 1 gives an overview of the total number of sources per country and their average trustworthiness scores (also see Fig. A.7 for the number of sources over time per country). NewsGuard also labels the language of a source, with English being the most represented (9,455 sources; 86.9%), followed by Italian (4.9%), French (4.4%), and German (3.8%).

Table 1: Country descriptives.

Country	n (%)	Trustworthiness (SD)	Updated	Changed	Stable since
US	8281 (76.1)	56.8 (34.1)	224 days	559 days	2021-03
GB	568 (5.2)	78.5 (23.5)	224 days	474 days	—
IT	537 (4.9)	70.9 (24.2)	291 days	661 days	2020-05
CA	471 (4.3)	87.7 (15.3)	306 days	430 days	2022-04
FR	424 (3.9)	65.4 (29.2)	326 days	646 days	2020-06
DE	368 (3.4)	77.0 (31.2)	308 days	673 days	2020-12
AU	166 (1.5)	72.5 (24.6)	331 days	358 days	—
AT	42 (0.4)	77.2 (29.6)	338 days	375 days	—
NZ	22 (0.2)	76.3 (24.7)	313 days	426 days	2024-01

Note. As of September 15th, 2024. Sorted by number of sources.

Across all versions of the database and in most countries, the overall trustworthiness of sources remains above 65, except for the United States. US-based outlets are of lower trustworthi-

ness on average (see Table 1). Figure 2C shows the average trustworthiness of sources per country over time (see also Figures A.8 and A.7 for plots by country). The trustworthiness score for US sources was lower from the first database version onward and remained consistently lower compared to other countries when considering the number of sources rated at comparable points in time. For example, the average score for US sources was lower six or twelve months after their addition to the database compared to French sources at the same interval. This suggests that US sources may generally have lower trustworthiness overall, rather than the finding being solely explained by a larger coverage of low-trustworthy outlets.

When comparing the number of sources that meet the journalistic criteria per country, we observe some striking patterns. Figure 2B shows the percentage of sources that *do not fulfill* a given indicator by country, i.e., a high percentage indicates that many sources fail to meet the criterion. For most countries, more than 50% of sources meet each criterion. Interestingly, over 70% of Italian sources do not provide names of content creators, and over 80% do not effectively correct errors. A similar but less pronounced pattern is visible for French sources. Across all criteria, the United States shows the highest percentage of sources not fulfilling the criteria, with 56.4% of sources failing to gather and present information responsibly, almost twice the share of sources than in any other country. Furthermore, 30% of US-based sources fail to fulfill the heavily-weighted criterion “Does not repeatedly publish false or egregiously misleading content”.

Intervals between updates of sources vary by country (see Table 1), ranging from 224 days for the US and 291 days for Italy to 308 days for Germany. Sources get updated once a year per country, on average. Overall, the US coverage seems the most stable and complete. However, the raw number of media outlets in each country may depend on numerous factors (e.g., population size, fragmentation of the media system, political system, and so on) and should therefore not be used as an indicator for good coverage alone.

Temporal analysis of country coverage. Overall, the development of trustworthiness indicates that for countries with a higher number of sources, the database seems to have settled into a stable state. In contrast, for newer countries (Canada, New Zealand, and Australia), rating coverage still seems to be patchy. To estimate when a country achieves a stable coverage state, we determine a window of “poor coverage” per country. To this end, we developed two criteria to evaluate stability, visualize the results in Figure A.7, and list the first date on which a country reached stable

coverage in Table 1. First, we examine the development of average trustworthiness scores over time as a proxy for a potentially biased inclusion of more untrustworthy (or trustworthy) sources when a country is first added to the database. We generally observe higher variability in the first months up to a year after a country is added to the database. Therefore, we assess the stability of the overall trustworthiness for a given country in 6-month intervals. Within every interval, we calculate the difference in average trustworthiness from one month to the next. We then compare this difference to the overall variability of trustworthiness in the country, using a conservative threshold of half the standard deviation of trustworthiness over the entire observation period. If one month-to-month change in a given 6-month interval surpasses this overall variability threshold, we classify the entire interval as unstable with respect to overall trustworthiness.

Second, we analyze the growth in the number of sources included in the database for a given country. While the total number of sources varies substantially between countries, the initial growth rate tends to be steeper. However, the number of sources continues to increase over time for all countries, which is why we calculate the percentage change in sources from month to month as a second stability indicator within 6-month intervals. Here, we again apply a conservative threshold of 5% growth. Therefore, similar to the stability criterion relating to the overall trustworthiness, if in a given month a number of sources larger than 5% of the already existing number of sources in the database was added, the entire 6-month interval is considered unstable with respect to source additions.

We define a country as having reached stable coverage if two consecutive 6-month periods (one year) meet both stability criteria. According to our criteria, this means that the sources added during a stable period are likely not systematically biased towards more trustworthy or untrustworthy sources and that no major news outlets are missing. We visualize the development of trustworthiness and source coverage for each country in Figure A.7 with stable windows shaded in gray. Our analysis shows that coverage of the US, Italy, France, Germany, and Canada has been stable since 2022. Further, New Zealand is stable from 2024 onward and Austria and Australia show a tendency to stabilize in the next 6-month time interval. While Great Britain has instability in the number of sources throughout, the trustworthiness score has been relatively stable since early 2022. To better understand how complete the database is per country, we describe comparisons with domains shared on social media and with other lists in the following sections.

Comparison with domains shared on social media. In the supplementary material of Lasser et al. (2022), the authors describe a manual inspection of web domains in their Twitter (now X) dataset that were not covered by NewsGuard (see “Reproducibility of downstream research results using different snapshots of the database”). More precisely, the authors investigate all links posted by politicians that point to web pages other than social media platforms and search platforms. Overall, NewsGuard seems to cover a notable fraction of those links (US: 46.5%, Germany: 58.8%, UK: 39.2%). Per country, the authors selected relevant domains not covered by NewsGuard that individually accounted for at least 0.1% of all linked domains. Subsequently, they inspected the missing domains and manually assigned labels to them (e.g., “government”, “news”, or “blog”).

For the US, 47 domains were analyzed, accounting for 21.2% of missing domains in the dataset. While most of these domains link to government or personal websites, none of the 47 domains lead to a news site. For the UK, out of the 78 analyzed domains (40.3% of missing domains), four domains not covered by NewsGuard are news sites, comprising a total of 1.5% of the links shared by the politicians. Of these sites, one was a clearly right-leaning site and two were mostly linked to by members of the labour party, while the last site was a news aggregator with no clear political leaning. Similarly, for Germany only a fraction of links not covered by NewsGuard lead to news sites (1.2%), corresponding to four out of the 62 manually analyzed missing domains (31.5% of missing domains). These sites were primarily shared by members of the left, green, or liberal parties, indicating that they are not widely shared across party affiliations.

Although the investigation examines only 20-40% of domains not evaluated by NewsGuard, by analyzing the domains that were most frequently shared, it demonstrates that NewsGuard provides a comprehensive list of domains shared by politicians in their social media communication across the US, UK, and Germany. Moreover, the small number of missing news sites and the relative absence of sources from both the left and right political spectrum in the missing domains suggests that NewsGuard’s domain curation shows no systematic political bias.

Comparison with other lists. In order to assess the actual coverage of sources per country, we compare the sources covered by NewsGuard with other lists. However, other lists are often not compiled with the same objective as NewsGuard, which is to cover all sources responsible for most of the internet traffic. Therefore, their scope and rating system may differ. In addition, other lists are not available for every country.

Primarily for the US context, Lin et al. (2023) have compared the overlap and agreement between ratings across six lists of existing expert ratings, including the NewsGuard database. While trustworthiness ratings correlated significantly for the web domains included in multiple lists, the authors found a high number of non-overlapping web domains between lists. For the German context, Puschmann et al. (2023) collected a list of 1,147 unique German online news domains (GOND)⁶, largely relying on web tracking data of a representative sample of 1,500 German citizens. Overall, GOND has 552 sources in common with NewsGuard (see Table A.2, among others labeled as “legacy press” (52.9%), “digital-born news outlet” (12.0%), or “hyperpartisan news” (10.1%). NewsGuard covers 244 out of 573 German-speaking sources covered by GOND, including some other languages.

Overall, comparisons of NewsGuard with the two lists for the US and Germany reveal many non-overlapping domains. However, NewsGuard’s coverage of the United States and Germany seems extensive in that NewsGuard generally tends to cover more prominent outlets and excludes niche and regional outlets, probably because it focuses on websites with the most engagement.

How valuable are contextual source labels for misinformation research?

Political orientation. In addition to trustworthiness ratings, the NewsGuard database also classifies sources according to their political orientation, relying on a left-to-right conception of the political spectrum. In December 2022, the categories were condensed from four (far left/slightly left/slightly right/far right) to two (left/right; see Fig. 4A). Overall, the political orientation label is only available for a minority of the sources, specifically 33.4% (3,789 sources) in the current database version.

In the US, 34.8% of news outlets have a political orientation label, in contrast to only a small minority of sources in other countries: 10.4% of sources for Great Britain, 13.2% for Italy, 16.9% for Germany, and 20.3% for France. Figure 4C and Table A.3 show the total number of political orientation labels per country as well as the number and percentage of sources classified as left or right-leaning in the most recent database. In most countries, the majority of sources with political orientation labels belong to the right-wing spectrum. Figure A.9 shows the distribution of political orientation per country over time.

⁶<https://osf.io/s5uhb/>

Trustworthiness ratings are generally lower for right-leaning sources (see Fig. 4B). In the database snapshot from September 2024, the average trustworthiness score for right-leaning sources is 26.4 ($SD=20.3$), in contrast to 60.6 ($SD=26.1$) for left-leaning sources ($M_{diff}=34.2$, $t(3771)=35.4$, $p<.001$). When political orientation was still divided into four categories, slightly left- and right-leaning sources had moderate trustworthiness ratings (for example, in June 2021, 81.3 and 59.4, respectively), and far-left and far-right sources had comparatively lower ratings (54.0 and 15.9, respectively). In August 2021, however, when a number of untrustworthy sources ($M=40.5$, $SD=37.2$) were added to the database, the average trustworthiness of sources categorized as slightly right-leaning dropped by almost 20 points. Figure 4A shows that the “Slightly Right” category dropped below the “Far Left” category by almost 15 points.

In addition to the change in average trustworthiness scores within political orientation categories caused by the changed taxonomy of political orientation used by NewsGuard, the trustworthiness averages per political orientation also differ by country, as shown in Figure 4B. While for some countries like France and Canada, the average trustworthiness of left and right sources only differs by about 20 points, other countries like Australia, Germany, and Austria show extreme differences of over 40 points.

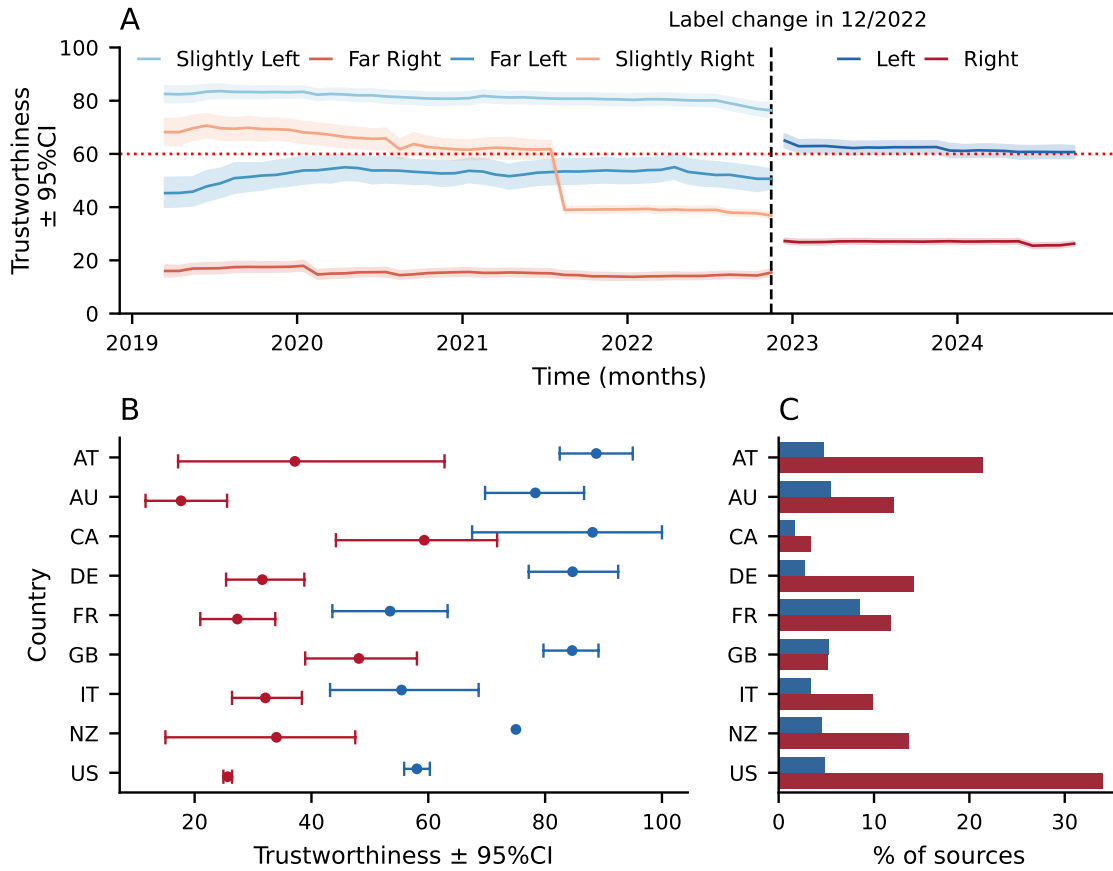


Figure 4. Trustworthiness averages by political orientation. Panel A: Over time (with 95% confidence intervals). Panel B: By country (with 95% CIs). Panel C: Percentage of sources per country with political orientation, explaining differences in trustworthiness by country.

Nevertheless, across all countries, sources classified as right, on average, always have a lower trustworthiness score than sources on the left. However, given the small percentage of sources classified as belonging to either side of the spectrum, we cannot infer that sources on the right, in general, have lower trustworthiness.

To assess whether sources with labels are correctly classified and whether sources without labels are neutral or not classified by NewsGuard for other reasons, we compared NewsGuard ratings with manual and other expert political source labels. First, we manually annotated the political orientation of all German-speaking sources ever covered by NewsGuard (as of September 15th, 2024) based on the website's front page, article headlines, ads, and article content. We followed the labels by NewsGuard (right/left) and added two labels: neutral and unidentifiable (e.g., when the website is unreachable). For sources with available ratings, the agreement between NewsGuard and the manual annotations was 83% (for 64 domains out of 406, excluding 20 dead links). Second, we compared the NewsGuard labels to labels by MediaBiasFactCheck⁷ (MBFC, 41 labels in total, six overlapping with NewsGuard), resulting in an agreement of 83%. Both comparisons showed that disagreements mainly occurred in cases where the manual annotations were labeled as neutral, while the expert ratings assigned a specific label. In only one case, NewsGuard differed from both MBFC and the manual annotation.

Consequently, in the German context, the existing ratings largely matched manual political orientation labels. Interestingly, the rated sources tend to be niche outlets with more extreme (and thus, clear-cut) political leanings. For the neutral sources, we could not identify a clear political leaning. Such sources were typically local newspapers or magazines focusing on special interest topics (e.g., technology or health). However, roughly half were legacy media or major online sources that slightly leaned towards a particular political orientation. Our findings suggest two (not necessarily mutually exclusive) implications for a missing political orientation label: Either the sources have no overt political bias in their reporting or are too niche to be labeled by NewsGuard.

Before the rating became binary, a few domains frequently oscillated between more extreme and less extreme labels, but domains rarely switched their political orientation. This might have been a reason for simplifying the label. Only five domains have changed their political orientation rating from left to right (none have changed from right to left). These domains are US-based, and two of them also experienced a significant drop in their trustworthiness score, reducing it by 42 points in January 2020. We therefore conclude that political orientation labels generally remain stable for the relatively small number of sources that have such a label.

In summary, although the political orientation labels of sources provided by NewsGuard are

⁷See here for rating system as of August 20th, 2024: [cd https://mediabiasfactcheck.com/methodology/](https://mediabiasfactcheck.com/methodology/)

sparse, particularly outside the US, the ratings of covered sources appear generally stable over time. An in-depth investigation of labels for German sources showed that NewsGuard’s political orientation assessment is reasonable and the low number of labels can be explained by sources not showing a political leaning or being too niche. Therefore, the political orientation rating by NewsGuard may be a valuable addition to analyses related to source trustworthiness. However, we recommend that researchers (at least) spot-check the ratings for the specific country in which they intend to use the database, especially for major outlets and mainstream media sources. If the research focuses on comparing political bias rather than controlling for it, we advise validating and extending the existing ratings provided by NewsGuard.

Topics. NewsGuard experts assign labels for the topics covered by the sources. Usually, a single source has multiple assigned topics. The number of sources with topic labels has steadily increased since their introduction in October 2019. As of May 2021, roughly 50% of the sources had topic labels (see Fig. A.10) for the total number and proportion of sources with topic labels). In September 2024, only 352 sources are without a topic label (2.9%). In Figure 5A, we show the distribution of trustworthiness ratings within topics while in Figure 5B, we show raw topic label counts. Note that a single source can have more than one topic label. The most popular topics are “Local news” (40.3% of sources), “Political news or commentary” (39.5%), “General news” (22.7%), “Health or medical information” (13.6%), and “Conspiracy theories or hoaxes” (12.8%). Over time, topic labels related to misinformation have been added, most likely coinciding with some of the major database updates discussed above. Figure 5C shows the top five topics discussed by sources classified as untrustworthy (measured by the percentage of untrustworthy sources publishing on those topics). In April 2020, shortly after the COVID-19 pandemic started, NewsGuard increasingly added a label called “COVID-19 misinformation”, with major additions in August 2021. Once a source receives a topic label, it rarely changes (except for foxnews.com, which has changed seven times).

As shown in Figure 5A, average trustworthiness ratings greatly differ across topics (white dots show the average trustworthiness of sources covering that topic). The lowest average rating is given to sources covering issues related to conspiracy theories ($M=12.6$, $SD=12.6$), military ($M=31.8$, $SD=34.9$), and health or medical information ($M=39.3$, $SD=35.5$). In contrast, those covering education have the highest average trustworthiness scores ($M=85.8$, $SD=16.9$), but this is only 0.5% of the sources. Generally, less prevalent topics tend to have high trustworthiness scores

with low variance (e.g., fashion with 76.5 on average, $SD=10.2$), see also Figure 5B. Considering only untrustworthy sources, however, only 32.5% are labeled as covering conspiracy theories. Instead, untrustworthy sources seem to commonly publish “Political news or commentary”, with 79% of all untrustworthy sources (e.g., with a trustworthiness score < 60) covering this topic, followed by local (30.4%) and health news (22.3%). These topics generally receive a significantly higher number of labels from NewsGuard and show a high variance in trustworthiness, for instance, “General news” or “Local news” (see Fig. 5A and B). This highlights that both untrustworthy and trustworthy sources cover a wide range of topics, including mainstream ones, and that fringe topics only make up a small proportion of the news.

Across countries, sources more or less publish on the same topics (see Fig. A.11 in the appendix). However, sources with low trustworthiness seem to publish on different topics, i.e., the trustworthiness of sources covering a given topic substantially depends on the country (see Fig. A.12). For instance, for medical information, sources in Canada are rated as trustworthy ($M=65.4$, $SD=34$, 27 sources), while trustworthiness is well below 60 in the United States ($M=30.2$, $SD=33.3$, 1,130 sources). The difference between countries is also substantial for political news, with an average trustworthiness of 33.1 ($SD=24.8$, 3,498 sources) in the US, 39.2 ($SD=29.3$, 111 sources) in France, and 81.6 ($SD=18.6$, 268 sources) in Great Britain.

We also analyzed the interaction of average trustworthiness by topic and political orientation (see Figure 5A). Political and local news are among the most popular topics assigned to both left- and right-wing sources, with 386 and 113 as well as 2,946 and 1,140 sources covering those topics, respectively. These numbers show the differences in how many political orientation labels are assigned to each side. For sources classified as right-wing, trustworthiness scores for all topics are consistently lower on average than for sources on the left side, e.g., political news: $M=26.6$ ($SD=20.0$) on the right and $M=60.2$ ($SD=26.0$) on the left. 1,162 right-wing sources cover conspiracy theories ($M=10.0$, $SD=11.4$) and 693 sources cover health-related news ($M=11.3$, $SD=15.0$). Meanwhile, for sources labeled as left-leaning, trustworthiness is particularly low (below 50) when they cover the topics “Conspiracies”, “Military and defense” or “Health news”. However, due to the skewed distribution of political orientation in the database, it is impossible to say whether the above-mentioned differences in countries and political orientation are due to actual differences in the media landscape or sampling strategies.

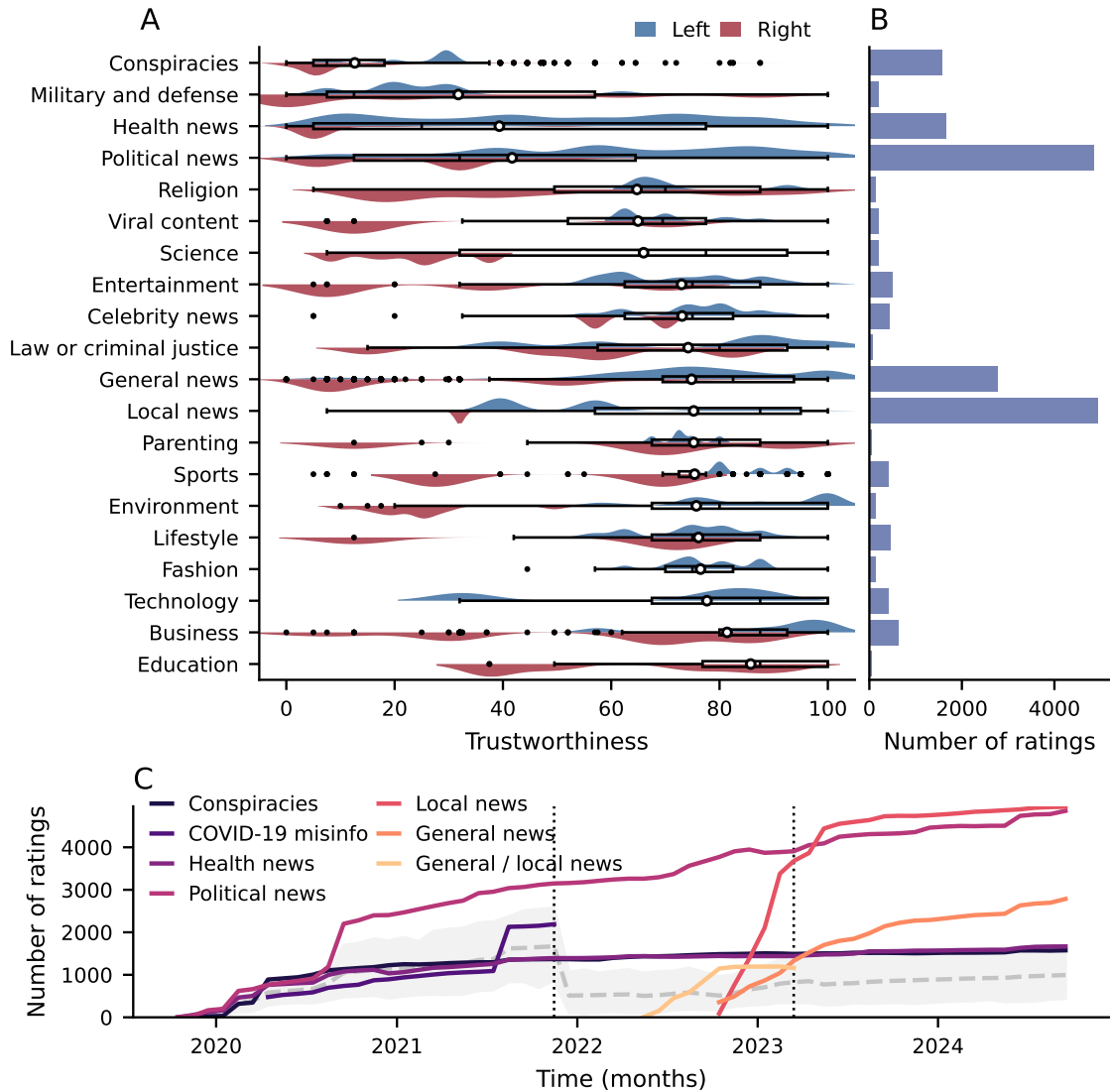


Figure 5. Panel A: Distributions of trustworthiness per topic and political orientation (left/right in blue/red, respectively), sorted by their average trustworthiness (white dot within the boxplot represents the quartiles and whiskers show the whole range). Note. We excluded topics with a count below 50. Some topics are abbreviated. Panel B: Count of topic labels as of September 15th, 2024. Panel C: Frequencies of misinformation-related labels over time. Colored lines show the top five topics covered by untrustworthy sources. The dotted, vertical lines show the removal of two of those labels. The gray and dashed line shows the average topic label count of those topics across all sources with the 95%CI.

Overall, the coverage of topic labels appears comprehensive, suggesting they could be a valuable asset for misinformation research. Especially a combination of trustworthiness and topic labels could complement source characterization, for instance, by distinguishing between fringe sources and those addressing mainstream issues. For sources that lack a label for political orientation, topics may help clarify if a source is genuinely neutral or tends to cover extreme topics low in trustworthiness.

Conclusion

An inherent challenge in misinformation research lies in how false information is defined and measured. A prominent approach for tracking online misinformation is to use a list of sources and their web domains. Due to its popularity and limited access, we examined the comprehensive NewsGuard database and analyzed the temporal stability and cross-country completeness of their trustworthiness ratings and other source labels relevant to misinformation research. Here, we summarize practical recommendations for using NewsGuard in research and discuss conclusions for source-based approaches in general.

Over 50% of the sources included in the database have a score of 60 points or higher, deemed at least “credible with exceptions” according to NewsGuard’s nomenclature. These sources typically meet two heavily weighted criteria: “Avoids deceptive headlines” and “Does not repeatedly publish false or egregiously misleading content”. The average trustworthiness score of sources included in the database has decreased over time due to the addition of untrustworthy sources rather than the degradation of trustworthiness of existing ones. NewsGuard provides a new snapshot of its database hourly, but an analysis of update time stamps of sources shows that source information is likely checked annually, with infrequent changes to trustworthiness ratings usually triggered by changes in transparency about website ownership and financing or editorial practices. Such changes are generally minor, involving updates of only one of the nine journalistic quality criteria – but may influence downstream research results if sources cross the binary classification threshold at 60 (“trustworthy”/“untrustworthy”). However, changes in source coverage can be abrupt, with instances of over 1,000 sources being added or removed in a single update. To shed some light on the impact of changes in the NewsGuard database on downstream research outcomes, we reproduced research investigating news-sharing practices of political elites in the US, Germany, and the UK (Lasser et al., 2022). We find that after 2019, the content of the database largely reached a stable

state in the US, and using different versions of the database between 2020 and 2024 does not change research outcomes if the continuous NewsGuard score is used. The process of stabilizing took longer for Germany, but after 2022, we also only observed minor database changes there. For the UK, the database never reaches a stable state according to our stability criteria, mainly due to the high frequency of abrupt changes in the number of sources included. However, the average trustworthiness ratings stabilize after 2022. In general, once a stable database state is reached, downstream research results relying on continuous ratings do not substantially change depending on the version of the database that is used.

Despite these nuances, the trustworthiness ratings by NewsGuard appear relatively insensitive to time, speaking for the reliability of source-based approaches. For countries and time periods where the database is stable, we also do not see much benefit for the use of dynamic scoring in which observations are matched to the corresponding snapshot of the database in time. As ratings of individual sources are relatively stable, we conclude that using a later version of the database with a higher coverage is preferable to using a version closer to the time of observation but with lower coverage.

The picture changes if binary “trustworthy” and “untrustworthy” labels for news sites are introduced by classifying sites based on a cutoff value at a NewsGuard score of 60. In this case, we observe large differences in the fraction of “untrustworthy” links that are shared by politicians, depending on the version of the database used for the analysis. We show that these differences can substantially change the direction and interpretation of research results in our replication study. As a result, we strongly caution researchers against using binary trustworthiness labels if continuous scores are available.

US-based sources generally score lower on trustworthiness than other countries; a trend that persists even when accounting for different points in time and the number of sources rated. Importantly, this does not say anything about the trustworthiness of news actually consumed by the US population, which could be higher, depending on the contribution of each source to the news diets of US Americans. Country-specific factors seem to significantly influence journalistic traditions across countries with similar media systems. Italy and France, part of the Southern cluster, show distinct patterns compared to Central (e.g., Germany) and Western clusters (e.g., US). According to Brüggemann et al. (2014), countries belonging to the Southern cluster have a less professional-

ized journalism sector than the Central or Western clusters. In their seminal definition of media systems, Hallin and Mancini (2004) state that in countries that fail to meet the professionalism indicators, the understanding of journalists serving the public interest is less pronounced. In other words, when journalistic professionalism is low, journalists lack professional autonomy, potentially explaining less institutionalized transparency practices in these countries, as indicated by NewsGuard's journalistic criteria. NewsGuard's criteria broadly reflect differences in media systems — at least for the Western, educated, industrialized, rich, and democratic (WEIRD) countries covered. Nonetheless, the differences in journalistic traditions underscore that applications of the database need to be tailored to individual contexts and that researchers should double-check the relevance of the criteria for the country studied⁸.

The NewsGuard database provides more than just trustworthiness ratings; it also includes labels for the political orientation of sources and covered topics, which could be valuable contextual information for misinformation research. While most sources have multiple topic labels, only a minority of sources have one for political orientation. About 41% of sources have a political orientation label in the US, but only around 10% do in other countries. In our manual analysis of German sources, we observed that NewsGuard tends to rate only sources with a clear political leaning. Most rated sources are right-leaning, and these are often deemed less trustworthy compared to left-leaning sources. This disparity raises the question of whether it stems from a sampling bias by NewsGuard editors rather than an actual difference in trustworthiness. We argue that NewsGuard seems unbiased for the countries where coverage and ratings have stabilized for the following reasons: First, NewsGuard selects sources based on web tracking data, suggesting that it should reflect the online media landscape for the respective country. Research on alternative media suggests the emergence of right-wing counter-publics as a response to the political and legacy media context (Heft et al., 2020). More precisely, the theory posits that a prevalence of left-leaning opinions in political and mainstream media spheres (Osmundsen et al., 2021) may lead to an under-representation and marginalization of right-wing views (Benkler et al., 2018) and fragmentation of a counter-public along the right-wing spectrum that is highly connected (Heft et al., 2021). As NewsGuard only labels the political orientation of explicitly partisan, sometimes hyper-partisan sources, this explanation would result in a higher number of smaller, right-wing online sources. A demand for right-wing fringe publications would also explain lower trustworthiness on the right-wing spectrum, which aligns with the political

⁸<https://www.newsguardtech.com/ratings/rating-process-criteria/>

asymmetry repeatedly observed in susceptibility to misinformation, e.g., recently by Robertson et al. (2023). Second, the source selection and trustworthiness ratings provided by NewsGuard are largely consistent with other independent fact-checking organizations (Lin et al., 2023). Therefore, it is unlikely that NewsGuard has an inherent bias against right-leaning sources, both in selecting more right-wing sources and in giving them lower trustworthiness ratings. NewsGuard also discusses this topic in their 2023 Social Impact Report⁹, emphasizing that their trustworthiness ratings are intended to be apolitical. Despite being incomplete, the political orientation labels may be a useful addition to the trustworthiness ratings, especially in combination with topic labels. However, they should be interpreted with caution and potentially validated and extended for countries other than the US and Germany.

Our analysis faces several limitations due to a scarcity of comparable datasets, particularly for countries other than the United States, the United Kingdom, and Germany. This limitation is especially critical when considering the absence of data on non-Western countries, which remains a substantial gap in the study of misinformation. Addressing this gap would require expanding the scope of databases like NewsGuard to include a broader range of countries, especially in the Global South, where the impact of misinformation may be profound but less studied. As a result of these limitations, we cannot assess potential biases in the selection of domains or the reliability and validity of the NewsGuard database globally; we can only evaluate the internal and temporal stability of existing ratings. To address this, we reproduced previous research findings using different database versions, compared the database with the few existing lists, and manually validated ratings. Based on these analyses, the NewsGuard database appears stable and complete for Great Britain, the US, Germany, and possibly France and Italy. Our analysis of temporal stability also shows that external events, such as changes in the media landscape or national and global crises, may disrupt stability in coverage. In particular, the onset of the COVID-19 pandemic appears to have caused instability in the US, Great Britain, and France during the early months of 2020. This can likely be traced back to efforts by NewsGuard to include ratings for sources relating to the spread of health (mis-)information that suddenly had much wider reach early in the pandemic. For countries where coverage has not stabilized yet, we recommend proceeding with caution and manually assessing potential biases in coverage that could influence research results by inspecting frequent domains that are not covered by NewsGuard. Additionally, our results are reproducible across different database versions, even with

⁹<https://web.archive.org/web/20240228003711/https://www.newsguardtech.com/special-reports/social-impact-report-2023/>

slight changes in the instrument, especially for its early versions. Given our observations, we recommend that researchers using NewsGuard's trustworthiness ratings examine several snapshots of the database to ensure no major source additions or deletions occurred recently. For countries included in NewsGuard for longer, e.g., the US, Canada, France, Italy, Germany and the UK, we suggest that researchers use whatever version of the database after 2022 they have access to. For more recent additions, e.g., Australia, Austria and potentially New Zealand, we suggest that researchers always use the most recent version of the database and carefully examine it for differences in coverage for different units of analysis, such as political leaning. Lastly, the topic and journalistic criteria ratings appear stable and may offer valuable context for the analysis of misinformation dynamics. A combination of political orientation and topic labels could effectively characterize sources and identify untrustworthy sources that address and decontextualize mainstream issues beyond hyper-partisan contexts. It is important to note that this analysis is specifically relevant to online misinformation research, as NewsGuard uses web traffic data to select sources for assessment. Future research should compare lists of online news domains to cross-platform and offline news consumption and examine how sampling based on sources versus stories influences downstream results.

In conclusion, relying on a list of rated online news sources offers a stable and scalable method for identifying content coming from untrustworthy online sources. A source-based approach with a fine-grained rating scheme based on journalistic quality criteria is particularly valuable and theoretically sound because it can encompass a wide range of sources, including less extreme forms of misinformation, thereby better-reflecting people's information diets. While not every article from an untrustworthy source is necessarily false – introducing some measurement inaccuracy – such sources often rely on sensationalist headlines or biased sourcing which diminishes their overall trustworthiness. Source-level ratings can reflect these broader patterns arising from weak editorial practices, even when individual articles vary in accuracy (Lazer et al., 2018). Here, the NewsGuard database serves as a useful tool for tracking online misinformation at the level of the source, but researchers should weigh its utility for the studied context against associated costs and consider alternative sources (e.g., Lin et al., 2023) for comprehensive analysis. Our findings underscore the critical need for dynamic, multifaceted, and openly accessible methods to get a clear and robust answer on the impact of misinformation on entire populations.

Acknowledgments

We acknowledge David Garcia and Noëlle Lebernegg for discussions about measurement issues in misinformation research and Annie Waldherr for her helpful feedback on the manuscript. J. Luehring, H.M., A.S. were supported by the Vienna Science and Technology Fund (WWTF) and the City of Vienna under grant 10.47379/ICT20028 and grant 10.47379/VRG16005. J. Lasser was supported by the Austrian Science Fund (FWF) under grant P 37280-N and the European Research Council (ERC) under the European Union's Horizon Europe program (Grant agreement No. 101160928). The funders played no role in study design, data collection and analysis, decision to publish or preparation of the article.

Data and Code Availability

Reproduction materials, including code and extended data, excluding the NewsGuard database, which is proprietary, are accessible on Github. To license the NewsGuard database, contact support@newsguardtech.com.

Author Contributions

Contributed to the conception and design: J. Luehring, H.M., J. Lasser, A.S.

Contributed to the analysis and interpretation of data: J. Luehring, R.M.L., J. Lasser, H.M.

Wrote the article: J. Luehring

Revised the article: J. Luehring, H.M., J. Lasser, A.S.

Final approval of the version to be published: J. Luehring, J. Lasser, H.M., A.S., R.M.L.

References

Allen, J., Howland, B., Mobius, M., Rothschild, D., and Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14):eaay3539. Publisher: American Association for the Advancement of Science.

Allen, J., Watts, D. J., and Rand, D. G. (2024). Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science*, 384(6699):eadk3451. Publisher: American Association for the Advancement of Science.

- Altay, S., Berriche, M., and Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media + Society*, 9(1):20563051221150412. Publisher: SAGE Publications Ltd.
- Aslett, K., Guess, A. M., Bonneau, R., Nagler, J., and Tucker, J. A. (2022). News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions. *Science Advances*, 8(18):eabl3844. Publisher: American Association for the Advancement of Science.
- Bakir, V. and McStay, A. (2018). Fake News and The Economy of Emotions. *Digital Journalism*, 6(2):154–175. Publisher: Routledge _eprint: <https://doi.org/10.1080/21670811.2017.1345645>.
- Baribi-Bartov, S., Swire-Thompson, B., and Grinberg, N. (2024). Supersharers of fake news on Twitter. *Science*, 384(6699):979–982. Publisher: American Association for the Advancement of Science.
- Benkler, Y., Faris, R., and Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.
- Bennett, W. L. and Livingston, S., editors (2020). *The Disinformation Age*. SSRC Anxieties of Democracy. Cambridge University Press, Cambridge.
- Bozarth, L., Saraf, A., and Budak, C. (2020). Higher Ground? How Groundtruth Labeling Impacts Our Understanding of Fake News about the 2016 U.S. Presidential Nominees. *Proceedings of the International AAAI Conference on Web and Social Media*, 14:48–59.
- Brüggemann, M., Engesser, S., Büchel, F., Humprecht, E., and Castro, L. (2014). Hallin and Mancini Revisited: Four Empirical Types of Western Media Systems: Hallin and Mancini Revisited. *Journal of Communication*, 64(6):1037–1065.
- Budak, C., Nyhan, B., Rothschild, D. M., Thorson, E., and Watts, D. J. (2024). Misunderstanding the harms of online misinformation. *Nature*, 630(8015):45–53. Publisher: Nature Publishing Group.

- Celadin, T., Capraro, V., Pennycook, G., and Rand, D. G. (2023). Displaying News Source Trustworthiness Ratings Reduces Sharing Intentions for False News Posts. *Journal of Online Trust and Safety*, 1(5). Number: 5.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., and Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29.
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., Tromble, R., Wilkins, A., Wojcieszak, M., de Jonge, C. K., Franco, A., Mason, W., Stroud, N. J., and Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656):392–398. Publisher: American Association for the Advancement of Science.
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425):374–378. Publisher: American Association for the Advancement of Science Section: Research Article.
- Guess, A., Aslett, K., Tucker, J., Bonneau, R., and Nagler, J. (2021). Cracking Open the News Feed: Exploring What U.S. Facebook Users See and Share with Large-Scale Platform Data. *Journal of Quantitative Description: Digital Media*, 1.
- Guess, A. M. and Lyons, B. (2020). Misinformation, disinformation, and online Propaganda. In Persily, N. and Tucker, J., editors, *Social media and democracy: The state of the field, Prospects for reform*, pages 10–33. Cambridge University Press, Cambridge.
- Guess, A. M., Nyhan, B., and Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, 4(5):472–480. Number: 5 Publisher: Nature Publishing Group.
- Hallin, D. C. and Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge University Press. Google-Books-ID: 954NJChZAGoC.
- Heft, A., Knüpfer, C., Reinhardt, S., and Mayerhöffer, E. (2021). Toward a Transnational Information Ecology on the Right? Hyperlink Networking among Right-Wing Digital News Sites in

- Europe and the United States. *The International Journal of Press/Politics*, 26(2):484–504. Publisher: SAGE Publications Inc.
- Heft, A., Mayerhöffer, E., Reinhardt, S., and Knüpfer, C. (2020). Beyond Breitbart: Comparing Right-Wing Digital News Infrastructures in Six Western Democracies. *Policy & Internet*, 12(1):20–45. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.219>.
- Lasser, J., Aroyehun, S. T., Carrella, F., Simchon, A., Garcia, D., and Lewandowsky, S. (2023). From alternative conceptions of honesty to alternative facts in communications by US politicians. *Nature Human Behaviour*, pages 1–12. Publisher: Nature Publishing Group.
- Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., and Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS Nexus*, 1(4):pgac186.
- Lazer, D., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., and Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380):1094–1096.
- Lin, H., Garro, H., Wernerfelt, N., Shore, J. C., Hughes, A., Deisenroth, D., Barr, N., Berinsky, A. J., Eckles, D., Pennycook, G., and Rand, D. (2024). Reducing misinformation sharing at scale using digital accuracy prompt ads.
- Lin, H., Lasser, J., Lewandowsky, S., Cole, R., Gully, A., Rand, D. G., and Pennycook, G. (2023). High level of correspondence across different news domain quality rating sets. *PNAS Nexus*, 2(9):pgad286.
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., and Hertwig, R. (2022). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, pages 1–28. Publisher: Nature Publishing Group.
- Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., and Nielsen, R. K. (2023). Reuters Institute Digital News Report 2023.
- Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing

- on Twitter. *American Political Science Review*, 115(3):999–1015. Publisher: Cambridge University Press.
- Pratelli, M., Petrocchi, M., Saracco, F., and De Nicola, R. (2023). Swinging in the States: Does disinformation on Twitter mirror the US presidential election system? arXiv:2303.12474 [cs].
- Puschmann, C., Stier, S., Merten, L., Kulshrestha, J., Rauxloh, H., and Schultz, C. (2023). German Online News Domains. Publisher: OSF.
- Robertson, R. E., Green, J., Ruck, D. J., Ognyanova, K., Wilson, C., and Lazer, D. (2023). Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature*, 618(7964):342–348.
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787. Publisher: Nature Publishing Group.
- Tay, L. Q., Lewandowsky, S., Hurlstone, M. J., Kurz, T., and Ecker, U. K. H. (2024). Thinking clearly about misinformation. *Communications Psychology*, 2(1):1–5. Number: 1 Publisher: Nature Publishing Group.
- Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., and Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44(2):157–173. Publisher: Routledge _eprint: <https://doi.org/10.1080/23808985.2020.1759443>.
- van der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, 28(3):460–467. Number: 3 Publisher: Nature Publishing Group.
- Vargo, C. J., Guo, L., and Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 20(5):2028–2049. Publisher: SAGE Publications.
- Weeks, B. E. and Gil de Zúñiga, H. (2021). What’s Next? Six Observations for the Future of Political Misinformation Research. *American Behavioral Scientist*, 65(2):277–289.

Yang, K.-C., Pierri, F., Hui, P.-M., Axelrod, D., Torres-Lugo, C., Bryden, J., and Menczer, F. (2021). The COVID-19 Infodemic: Twitter versus Facebook. *Big Data & Society*, 8(1):20539517211013861. Publisher: SAGE Publications Ltd.

Online Appendix

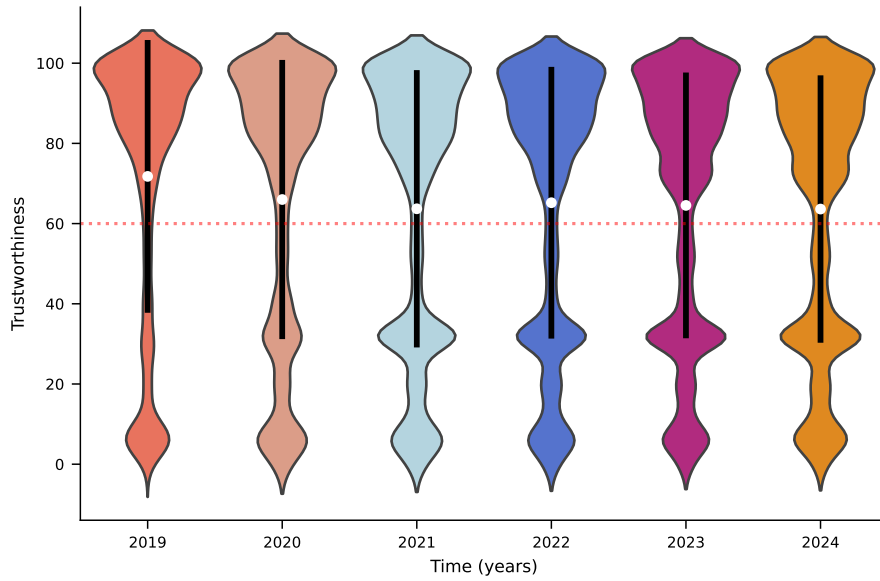


Figure A.1. Distribution of trustworthiness per year as a violin plot (with M and SD), standardized by count.

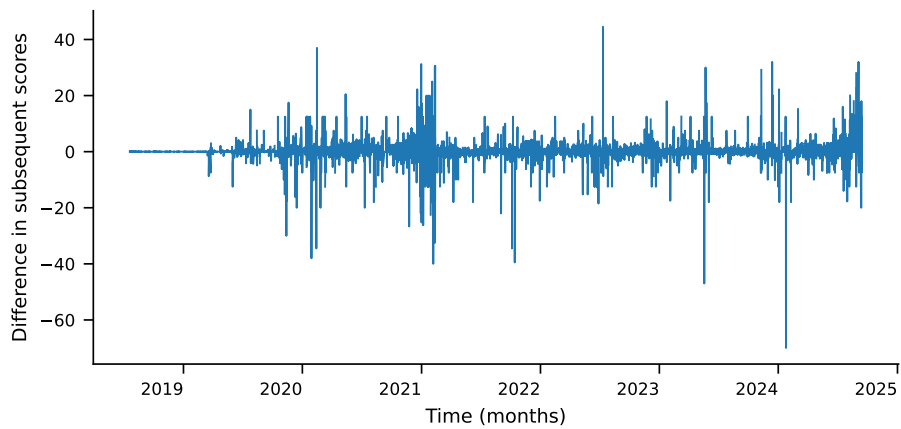


Figure A.2. Changes in trustworthiness scores over time, calculated as the difference in scores per news domain between consecutive updates.

Table A.1: Correlation matrix of journalistic quality criteria.

	1	2	3	4	5	6	7	8	9
1	1	0.63	0.4	0.56	0.93	0.38	0.39	0.26	0.36
2	0.63	1	0.57	0.75	0.63	0.58	0.49	0.55	0.56
3	0.4	0.57	1	0.48	0.4	0.45	0.44	0.34	0.53
4	0.56	0.75	0.48	1	0.57	0.52	0.51	0.47	0.5
5	0.93	0.63	0.4	0.57	1	0.39	0.37	0.28	0.39
6	0.38	0.58	0.45	0.52	0.39	1	0.42	0.48	0.4
7	0.39	0.49	0.44	0.51	0.37	0.42	1	0.29	0.36
8	0.26	0.55	0.34	0.47	0.28	0.48	0.29	1	0.47
9	0.36	0.56	0.53	0.5	0.39	0.4	0.36	0.47	1

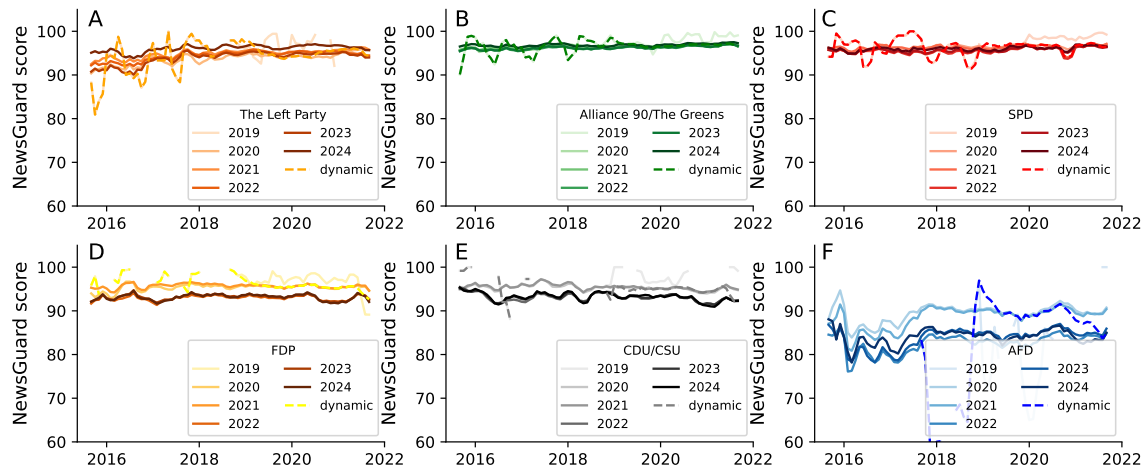


Figure A.3. Reproduction of trustworthiness ratings of sources shared by German politicians on Twitter over time from Lasser et al. (2022). Panel A shows average NewsGuard scores in Twitter posts by members of the party DIE LINKE, panel B for members of the Greens, panel C for members of the SPD, panel D for members of the FDP, panel E for members of the CDU and CSU, and panel F for members of the AFD.

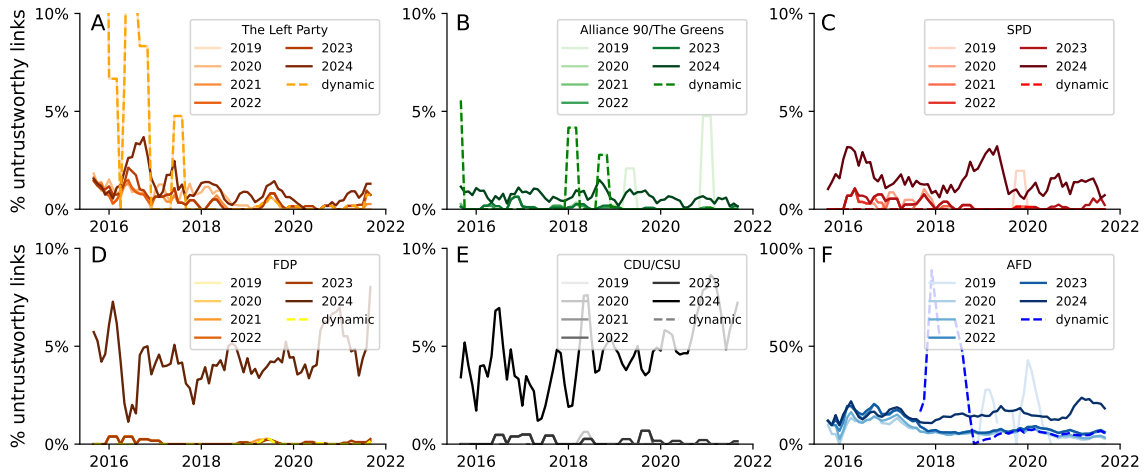


Figure A.4. Fraction of untrustworthy links (e.g., NewsGuard score <60) shared by German politicians on Twitter over time from Lasser et al. (2022). Panel A shows the fraction of untrustworthy links shared in Twitter posts by members of the party DIE LINKE, panel B for members of the Greens, panel C for members of the SPD, panel D for members of the FDP, panel E for members of the CDU and CSU, and panel F for members of the AFD—note the different scale in the y-axis!

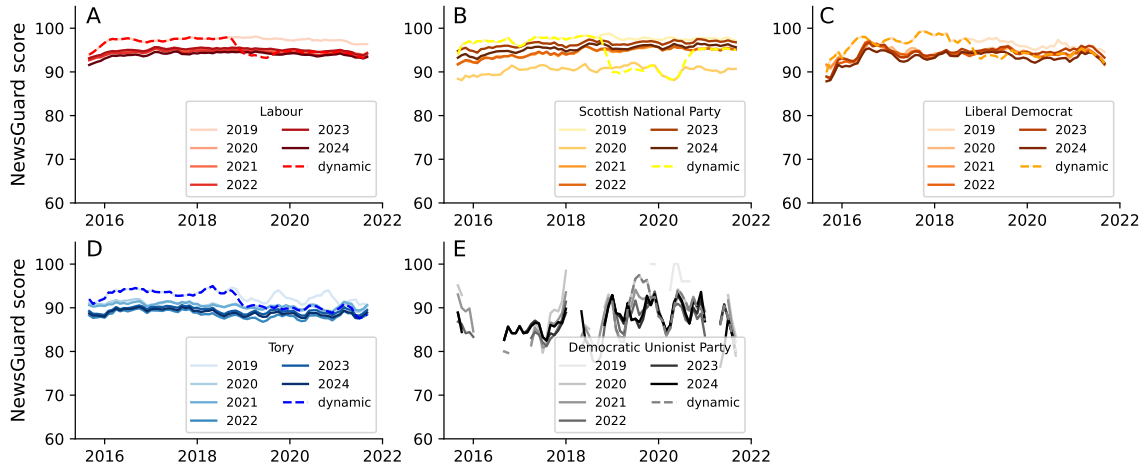


Figure A.5. Reproduction of trustworthiness ratings of sources shared by UK politicians on Twitter over time from Lasser et al. (2022). Panel A shows average NewsGuard scores in Twitter posts by members of the Labour party, panel B for members of the Scottish National Party, panel C for Liberal Democrats, panel D for Tories, and panel E for members of the Democratic Unionist Party.

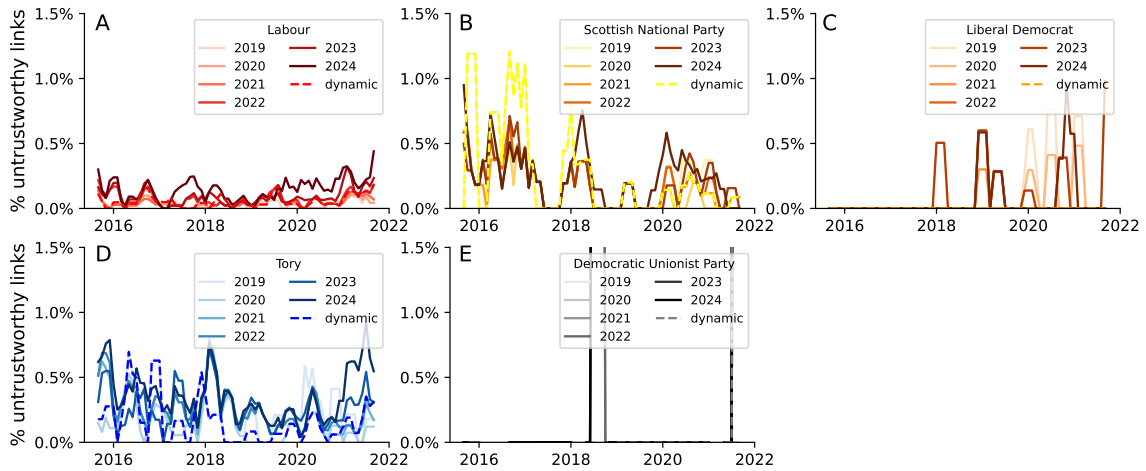


Figure A.6. Fraction of untrustworthy links (e.g., NewsGuard score <60) shared by UK politicians politicians on Twitter over time from Lasser et al. (2022). Panel A shows average NewsGuard scores in Twitter posts by members of the Labour party, panel B for members of the Scottish National Party, panel C for Liberal Democrats, panel D for Tories, and panel E for members of the Democratic Unionist Party.

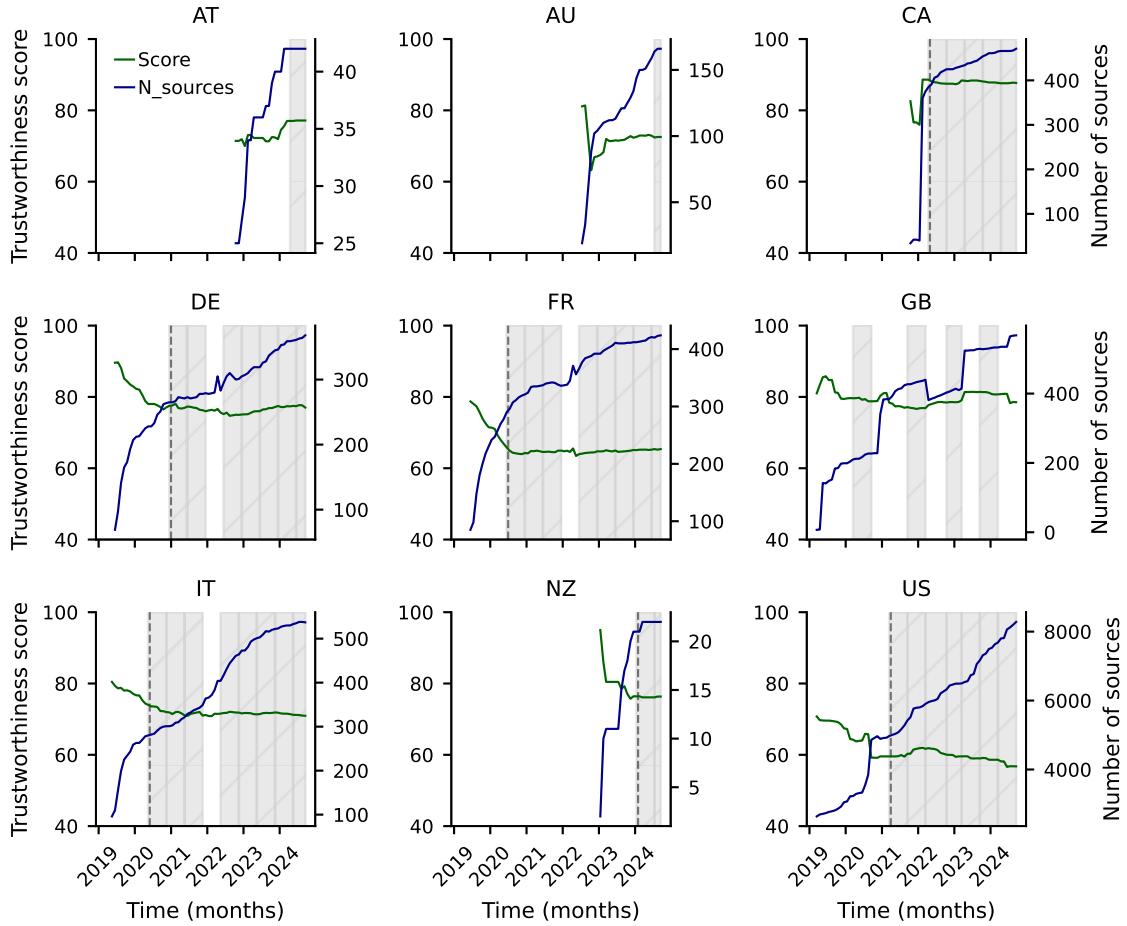


Figure A.7. Development of overall trustworthiness per country over time (green line) and the number of sources (blue line) with shaded areas representing stable 6-month windows. Dashed lines indicate the first date of a stable interval.

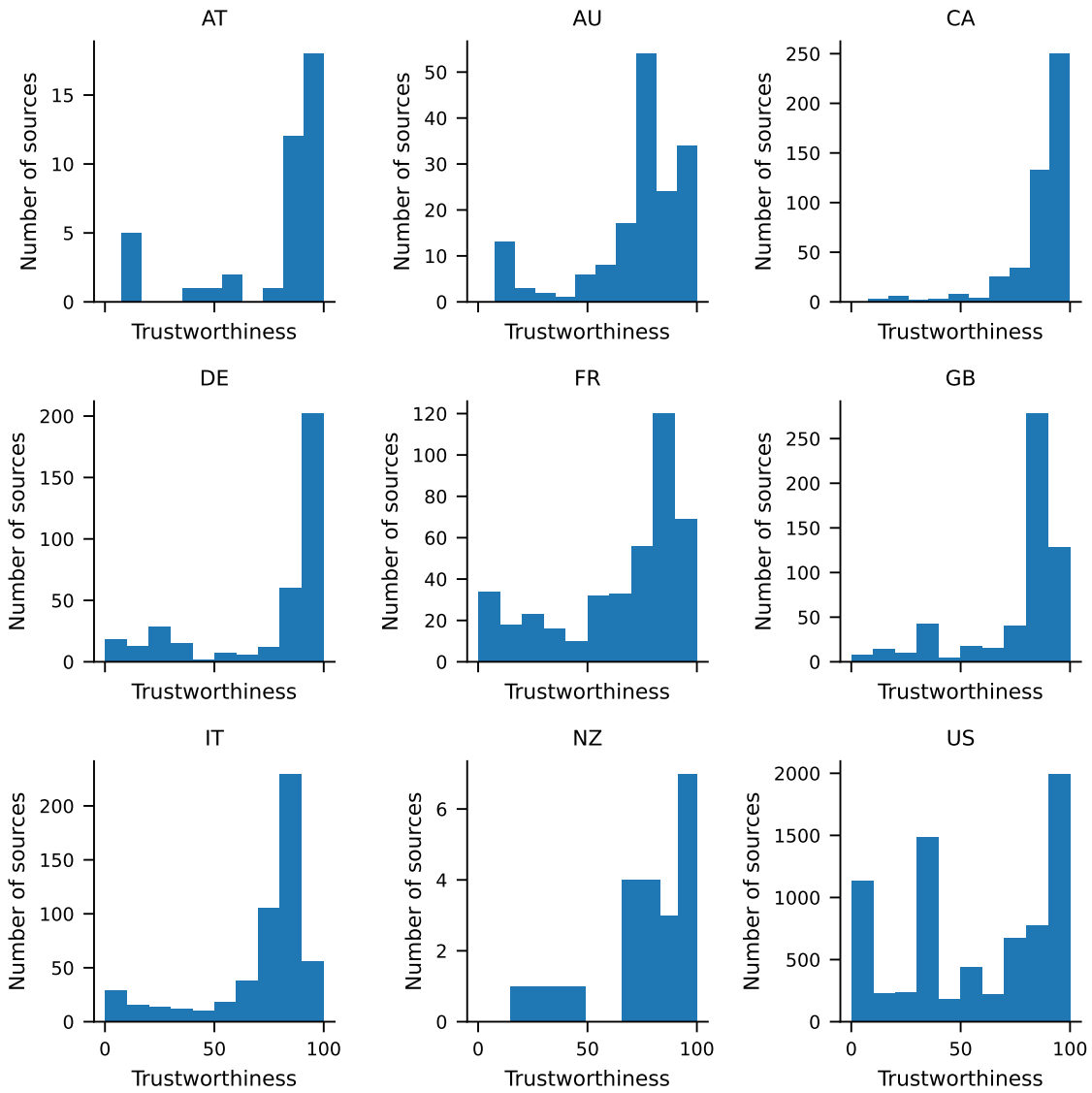


Figure A.8. Distribution of trustworthiness per country as of September 15th, 2024.

Table A.2: Overlap between GOND and NewsGuard.

GOND Type	NewsGuard Score	Overlap	%
commercial broadcaster	89.4	58	10.5
digital-born news outlet	84.5	66	12.0
hyperpartisan news	43.0	56	10.1
legacy press	91.0	292	52.9
public broadcaster	89.0	40	7.2
tabloid newspaper	84.6	40	7.2

Note. As of September 15th, 2024.

Table A.3: Orientation per country.

Country	Left (%)	Right (%)	Total
US	401 (12.45)	2819 (87.55)	3220
FR	36 (41.86)	50 (58.14)	86
IT	18 (25.35)	53 (74.65)	71
DE	10 (16.13)	52 (83.87)	62
GB	30 (50.85)	29 (49.15)	59
AU	9 (31.03)	20 (68.97)	29
CA	8 (33.33)	16 (66.67)	24
AT	2 (18.18)	9 (81.82)	11
NZ	1 (25.0)	3 (75.0)	4

Note. As of September 15h, 2024.

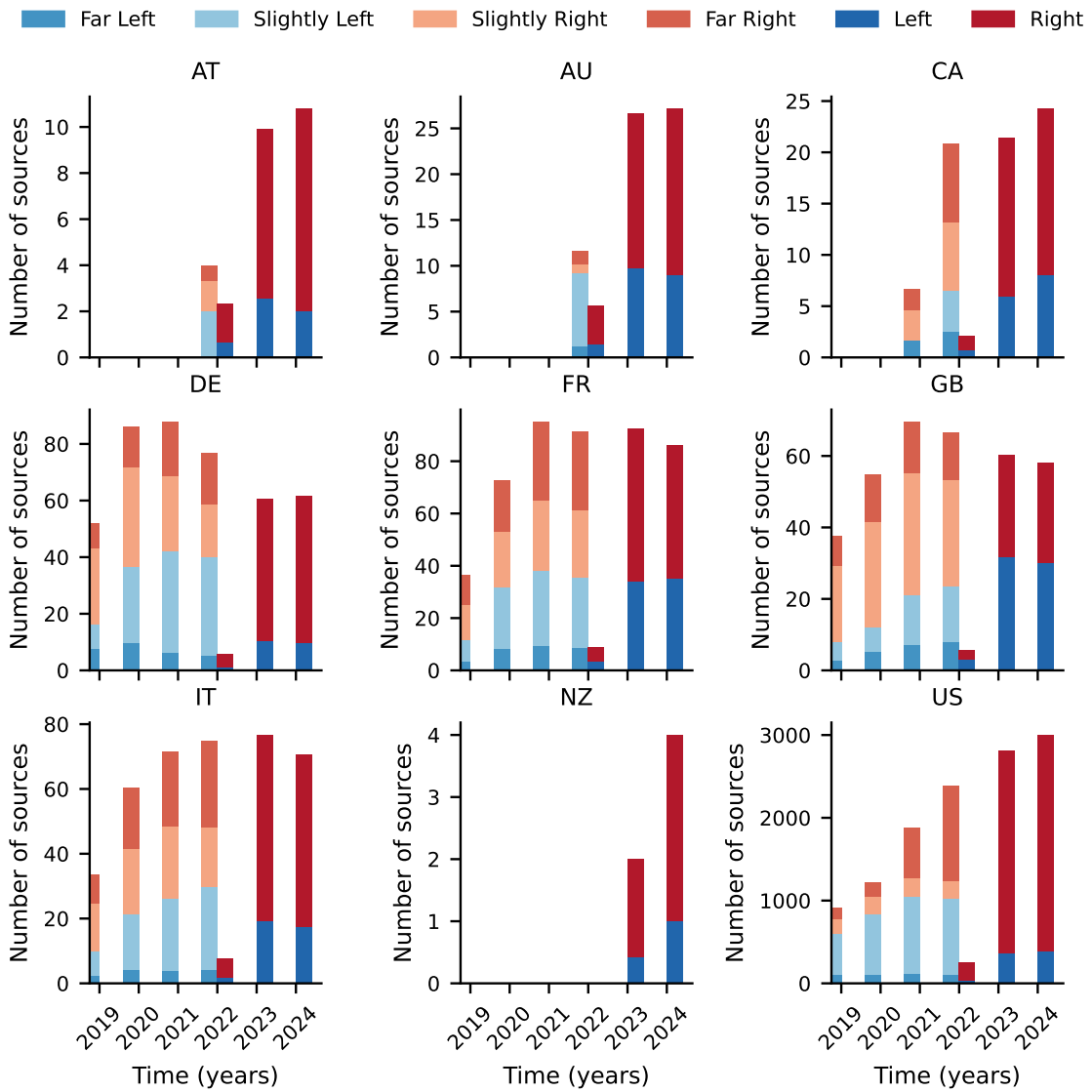


Figure A.9. Distribution of political orientation over time, by country

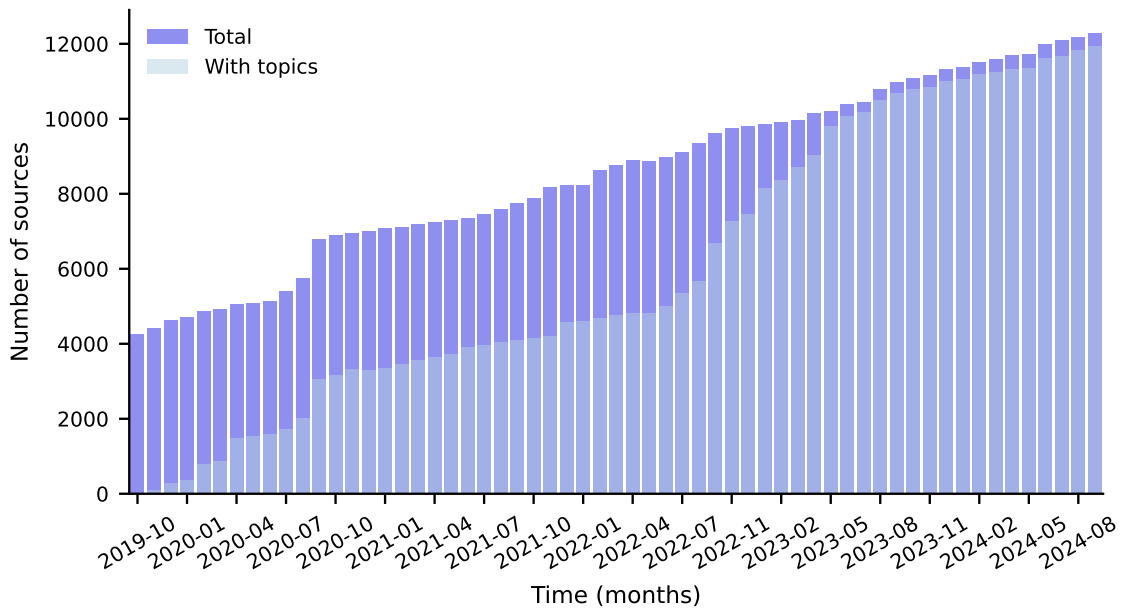


Figure A.10. Number of sources over time with and without topic labels.

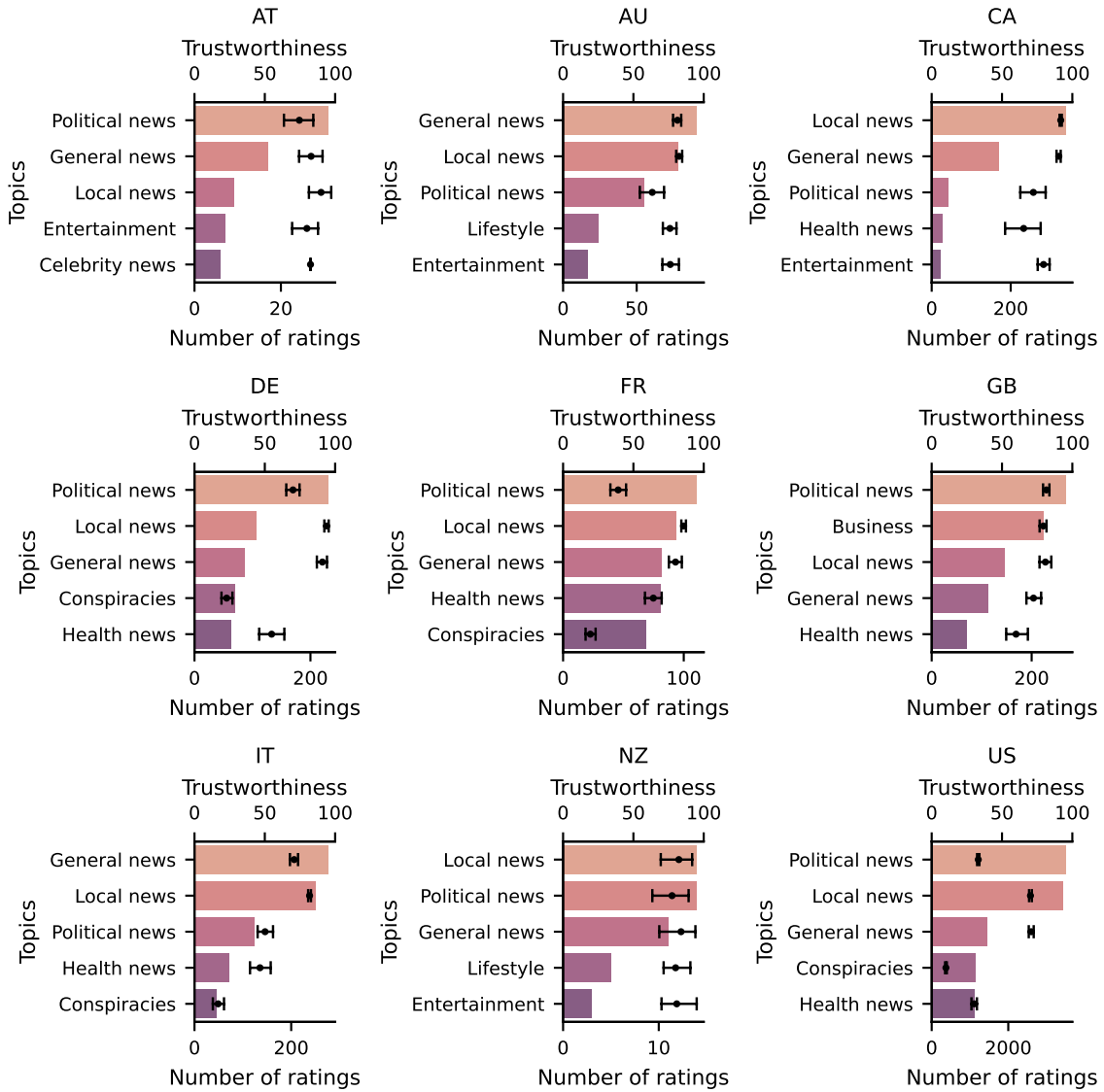


Figure A.11. Top five most popular topics per country and their total count, with the second x-axis at the top showing the average trustworthiness and standard deviations of those topics..

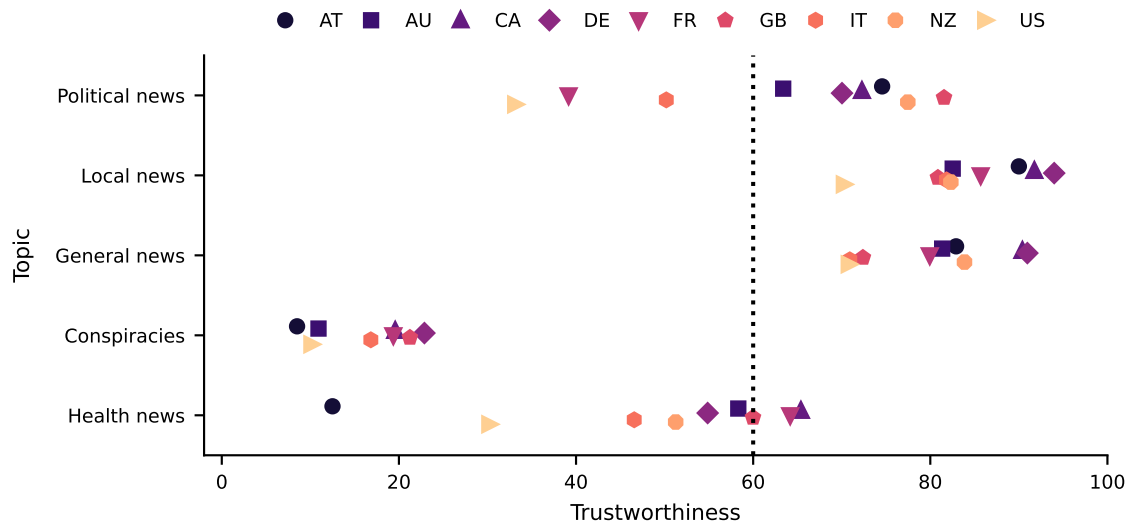


Figure A.12. Trustworthiness score for the top five most popular topics across countries in the most recent database (September 15, 2024).