nature human behaviour

Article

https://doi.org/10.1038/s41562-025-02223-4

Using co-sharing to identify use of mainstream news for promoting potentially misleading narratives

Received: 15 May 2024

Accepted: 23 April 2025

Published online: 10 June 2025

Check for updates

Pranav Goel[1] ✉, Jon Green[2], David Lazer[1,3] & Philip S. Resnik[4,5]

Much of the research quantifying volume and spread of online misinformation measures the construct at the source level, identifying a set of specific unreliable domains that account for a relatively small share of news consumption. This source-level dichotomy obscures the potential for users to repurpose factually true information from reliable sources to advance misleading narratives. We demonstrate this potentially far more prevalent form of misinformation by identifying articles from reliable sources that are frequently co-shared with (shared by users who also shared) 'fake' news on social media, and concurrently extracting narratives present in fake news content and claims fact checked as false. Specifically in this study, we use Twitter/X data from May 2018 to November 2021 matched to a US voter file. We find that narratives present in misinformation content are significantly more likely to occur in co-shared articles than in articles from the same reliable sources that are not co-shared, consistent with users using information from mainstream sources to enhance the credibility and reach of potentially misleading claims.

There is much concern in the social sciences about the prevalence of 'fake news', defined in ref. 1 as "fabricated information that mimics news media content in form but not in organizational process or intent", in the online information environment[2,3]. This 'process-driven' conceptualization of fake news has fostered 'source-level' measures of misinformation more broadly. Sources, and all their constituent stories, are categorized as reliable or unreliable largely on the basis of the respective sources' adherence (or lack thereof) to the procedures and norms of mainstream journalism, in addition to their reputations for veracity. These categorizations are then projected onto all of the information that the sources produce. This process is streamlined by the publication of domain lists, curated by researchers (for example, ref. 4) or third-party vendors (for example, ref. 5), that can be applied to any set of URLs for ease of classification[6–8].

Within this framework, information is evaluated not on the basis of its truth value, but instead on the reputation of its source. This provides a limited foundation for identifying and addressing the challenges misinformation poses. Practically speaking, 'fake' news sources capture a small share of overall news consumption[4,6,9,10]. For example, one study estimated that only 1% of individuals on Twitter accounted for 80% of all exposure to fake news sources[4], while another study concluded that over 90% of individuals on Facebook did not share any fake news sources[9]. If this constituted the sum of false or misleading content of empirical interest, then the widespread concern regarding misinformation[11–15] would seem to be misplaced.

One practical way in which domain lists might underestimate the extent of the misinformation challenge is these lists becoming quickly outdated as the information landscape evolves continuously[16]. However, the broader limitations of source-level measures of misinformation are theoretical. The binary categorization of sources as either 'fake' or 'reliable' obscures the reality that information produced by unreliable sources can be true, while information produced by reliable

[1]Network Science Institute, Northeastern University, Boston, MA, USA. [2]Department of Political Science, Duke University, Durham, NC, USA. [3]Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA. [4]Department of Linguistics, University of Maryland, College Park, MD, USA. [5]Institute for Advanced Computer Studies, University of Maryland, College Park, MD, USA. ✉e-mail: p.goel@northeastern.edu

https://doi.org/10.1038/s41562-025-02223-4

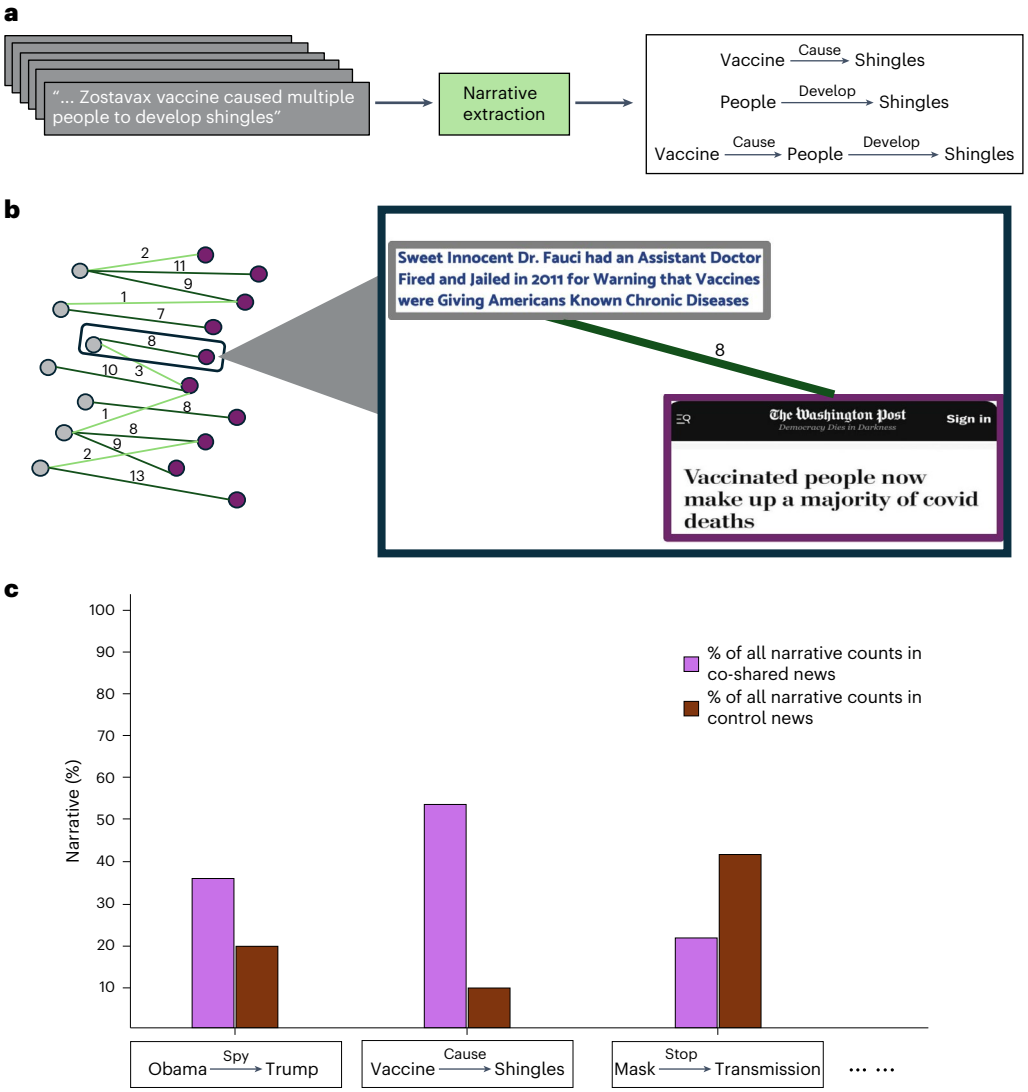

Fig. 1 | Visual illustration summarizing our approach. a, We automatically extract narrative structures from all texts in our data: misinformation content and mainstream news articles. Libraries of potentially misleading narratives are created using the narratives extracted from misinformation content. The narrative extraction approach is summarized in Fig. 3. **b**, To find reliable outlet articles disproportionately co-shared with fake news, we construct a graph with edges between fake and reliable articles (the nodes) weighted by the number of users sharing both incident articles. For example, the two articles highlighted above, one by a fake news outlet and one by The Washington Post, are both shared by eight unique individuals on Twitter. We then score all pairs of fake and reliable news on the likelihood of being shared by the same set of Twitter users, when controlling for the individual popularity of the two articles (edges in light green (or lighter) shade are below the threshold for being considered as co-shared). **c**, For each narrative from the potentially misleading narrative library, we compute their presence in co-shared news and non-co-shared (control) news. Presence is calculated as the percentage of articles in a set (co-shared or control) that a given narrative occurs in. Via an aggregated comparison across potentially misleading narratives, we find that these narratives present in misinformation content are significantly more likely to occur in co-shared articles than in articles from the same reliable sources that are not co-shared.

sources can be false or misleading. Moreover, the binary categorization of information as either true or false does not account for the extent to which information, especially on the internet, does not exist independently from its social function[17]. Groups of users do not share individual pieces of information for information's sake alone, but instead share information to support broader claims that advance their interests. These claims may be less true than the sum of their parts, as users may pick and choose true pieces of information from reliable sources to advance potentially misleading narratives. The extent to which information informs, rather than misinforms, often depends on how it is used.

We demonstrate this dynamic by examining the networked nature of information on social media. Specifically, we examine patterns in 'co-sharing' on Twitter by first identifying users who share information from unreliable sources and then examining the information from reliable sources that those users also share (while Twitter was rebranded as X in 2023, in this Article we continue to refer to the platform as Twitter since our dataset predates the rebranding). This allows us to score articles from mainstream sources on the basis of their propensity to be co-shared with 'fake' news. We then use automated tools to extract narrative structures from mainstream articles, 'fake' news articles and fact-checked statements known to be false. We expect that mainstream stories with high scores on this measure will be significantly more likely than those with low scores to contain narratives present in misinformation content, or potentially misleading narratives—entity–action relationships that repeatedly occur in information produced by unreliable sources or occur in statements known to be false. This suggests a dynamic in which users seeking to promote potentially misleading narratives may use factually true information to do so, to the extent to which mainstream sources produce

https://doi.org/10.1038/s41562-025-02223-4

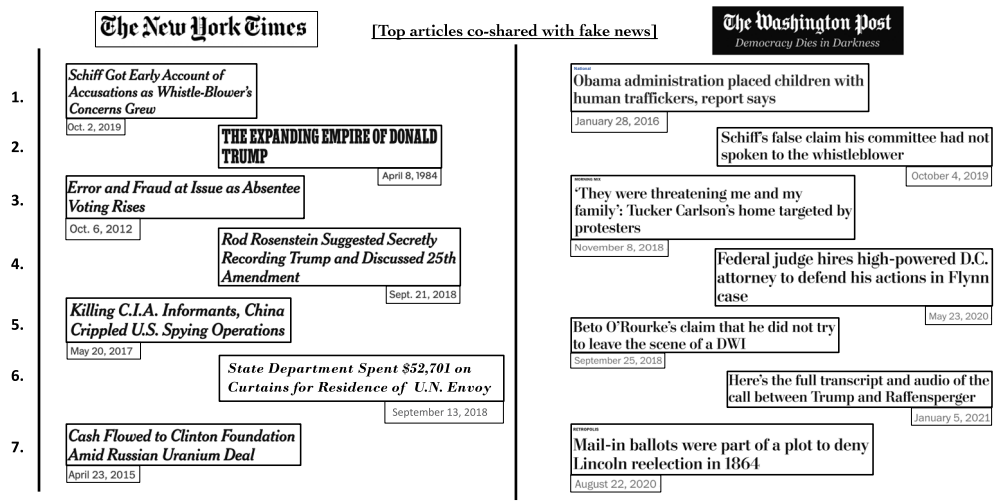

Fig. 2 | The top articles published by NYT and WAPO in the set of articles disproportionately co-shared with fake news. This is based on the co-sharing likelihood derived from Twitter sharing behaviour (specifically, based on an aggregated co-sharing likelihood score for each outward edge from a particular article or URL). As motivation for our hypothesis in this work, consider the third NYT article and the seventh WAPO article shown in the figure about mail-in ballots and absentee voting (which also pop up in our second qualitative case study). These articles are popular in the same Twitter circles sharing stories published by fake news outlets containing narratives (extracted in this work) that include "mail ballot raise risk fraud", "lot people cheat mail", "mail balloting increase incident fraud", "paper mail balloting mean dramatic increase fraud" and "republican complain potential mail absentee ballot fraud". While only headlines are shown here for illustrative purposes, our investigation also used the entire text of the news article. In this figure, we are only showing the top articles that are not opinion pieces, but we note that especially for WAPO, the top co-shared articles set does include opinion pieces.

information that is useful for this purpose. This dynamic reflects recent work showing that many partisan users selectively curate politically consistent information from politically diverse sources[18,19], challenging source-level measures of political slant. It further suggests that false and misleading claims might be far more prevalent on social media than source-level measures alone suggest, highlighting the importance of adopting a measurement approach that incorporates user behaviour and the interaction between information from different sources[20].

For example, consider an article (washingtonpost.com/politics/2022/11/23/vaccinated-people-now-make-up-majority-covid-deaths/) published by The Washington Post (WAPO) with the headline "Vaccinated people now make up a majority of covid deaths" (which has since been changed to "Covid is no longer mainly a pandemic of the unvaccinated. Here's why."). This article was published by a mainstream source and the headline is, strictly speaking, true. It was disproportionately shared by users in our data who 'also' shared anti-vaccine-related articles published by unreliable sources (Fig. 1b). Without important context—that since the vast majority of the country was vaccinated, the count of COVID-related deaths would be higher among the vaccinated even if the infection 'rate' remained much lower—one can see how this content would be useful for a user who wished to promote the false narrative that the COVID-19 vaccine is ineffective or actively harmful. This narrative is popular in COVID-19-related misinformation content[21] and is consistent with work showing how vaccine scepticism can be promoted through misleading content rather than direct falsehoods[22]. As this illustrative example suggests, those who wish to promote potentially misleading narratives such as these can strategically repurpose factually correct information to do so, when such information is available.

We test our expectations by deploying a tool for automatically extracting narrative structures from English text that are represented using narrative strings or labels. This provides us with narratives present in fake news articles as well as claims fact checked to be false (Fig. 1a). Our experiments across a variety of collections of potentially misleading narratives and types of reliable outlet support our primary hypothesis (the testing procedure is visualized in Fig. 1c). The difference in the presence of potentially misleading narrative strings between co-shared and 'control' articles (those published by the same set of mainstream news outlets but not co-shared with articles from unreliable sources) remains significant when accounting for partisan differences in the Twitter audiences of the two groups (Supplementary Text 2).

Our approach incorporates and systematizes conceptual frameworks in misinformation research that highlight users' own roles in advancing false and misleading claims[23,24]. Understanding the dynamics that advance these broader claims, and not just specific pieces of incorrect information, is critical—strictly true information repurposed to advance misleading claims can have harmful real-world consequences (such as reduced vaccine uptake)[22]. Methodologically, we provide a scalable and generalizable way of identifying misinformation that would not be captured by domain-level classification, complementing recent text-based approaches[25].

Finally, our findings have important implications for responsible journalistic practice. When vetting news stories (including the headlines) and their framing, it is important to consider not just whether the information contained in a headline or story is strictly true, but also whether it is likely to be 'used' in ways that are more broadly misinformative. We further highlight the possible consequences of framing choices in mainstream news and their potential to be repurposed for misinformation through two qualitative case studies, contextualizing how our framework can support journalism in practice. In particular, our qualitative case studies highlight how particular mainstream news articles co-shared with fake news can be repurposed to spread potentially misleading narratives: we uncover different types of co-shared mainstream article in terms of how they can be repurposed. These case studies cover two different topics: vaccine-related misinformation (of the form that imply vaccinated people spreading the disease or being the key node of transmission, especially in the context of COVID-19) and voter-fraud-related misinformation (allegations of voter fraud in the 2020 US election, particularly related to mail-in ballots). Both these topics of misinformation were prominent on Twitter and online social media in 2020[26–28].

Throughout this paper, we present examples of various kinds from our dataset to help develop a theoretical intuition for our research design and findings. Figure 1, as mentioned above, presents a visual overview of our entire approach, using an illustrative example.

https://doi.org/10.1038/s41562-025-02223-4

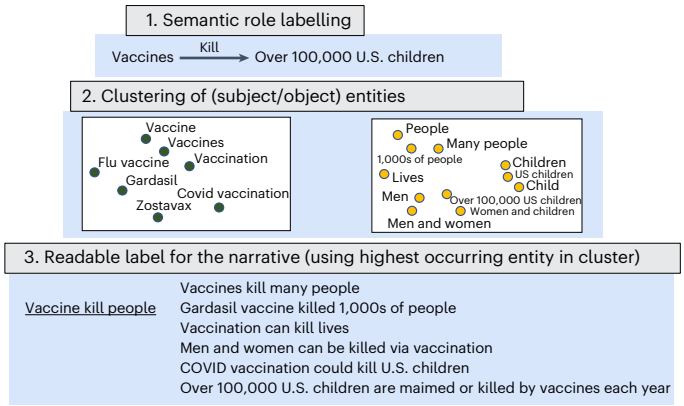

Fig. 3 | Brief summary of the process of extracting narratives. This uses the method created in previous work[30], and the visual is for the following sentence present in a fake news article in our dataset: "Over 100,000 US children are maimed or killed by vaccines each year".

Figure 2 provides the top co-shared articles from The New York Times (NYT) and WAPO. Supplementary Fig. 1 similarly presents the top co-shared articles from CNN and Fox News. Figure 3 provides a summary of the narrative extraction method using an example from our dataset. Table 1 provides the top 3 most frequent extracted narratives that are present in misinformation content as well as co-shared mainstream articles from reliable outlets but are absent in articles from the same reliable outlets that are not co-shared with 'fake' news. Our case studies are also a collection of various kinds of examples touching all parts of our dataset: we provide examples of many relevant narratives, selected mainstream news articles along with the fake news articles they were co-shared with, and examples of publicly available tweets (in our dataset) by Twitter users with a history of sharing relevant fake news, sharing selected co-shared mainstream articles—these examples provide valuable qualitative insights into 'how' mainstream news articles might be getting used for misleading purposes. These details and examples emergent in our case studies are provided in Supplementary Texts 3 and 4 and Supplementary Tables 5–25. The case studies also highlight a few of the mainstream articles that gain disproportionate popularity among fake news circles on Twitter in our dataset. Supplementary Tables 46–66 present all the narrative structures automatically extracted from claims fact checked as false.

In this paper, we use the terms 'domains', 'sources' and 'outlets' interchangeably to refer to the organizations that produce political information. To facilitate a clearer understanding of our methods and findings, we provide definitions and brief contextual backgrounds for the terminology used throughout this paper below:

'Narratives': These are sets of relationships or structures in language that describe how entities act on each other[29], characterized by a base component of the following format: AGENT ENTITY $\rightarrow^{\text{VERB}}$ PATIENT ENTITY[30]. Some examples are shown in Fig. 1a and Table 1; the extraction process is summarized in Fig. 3. This is a specific instantiation using some of the components that have been found to be crucial for characterizing narratives in text across the broad and varied literature on narratives (as well as timing, motivations and so on[31]).

'Misinformation': This broadly refers to false information. This often includes content published by outlets categorized as unreliable and specific claims fact checked as false. The former may also be referred to more specifically as 'fake' news, which is typically considered to be a subset of misinformation. Misinformation is also distinguished from *dis*information in that the latter involves an organized attempt to deceive.

'Fake news outlets' and 'fake news': Fake news outlets are sources that produce news via processes that do not conform to standard industry practices (including editorial norms and processes for vetting information)[1,4]. The definition we use here is closely related to the usage of 'the deliberate creation of pseudojournalistic disinformation' in previous work[32].

We classify outlets known to regularly publish false content as fake news sources, and articles published by those sources as fake news, using annotations from NewsGuard[5]. We aim for precision by considering 'regularly publishing false content' as our criterion out of all the factors that determine the trustworthiness score of an outlet in NewsGuard ratings. This helps us highlight the drawbacks of conceptualizing problems with the digital information ecosystem as that of 'false information' or fake news content—a framework that often foregoes closer examination of information coming from sources deemed reliable. Complete details about the use of NewsGuard annotations in our work, description of NewsGuard and their methodology, and the rationale for using these annotations are provided in Methods.

'Reliable outlets': These are all news outlets annotated for their quality as information sources by journalists at NewsGuard and 'not' deemed as fake news outlets according to our selection criteria above. Since this is quite broad, we also show our primary results with narrower criteria such as only considering sources marked as 'trustworthy' by NewsGuard. as well as partisanship-based subsets (details provided in Methods).

'Mainstream media or news outlets': Within reliable outlets, we define mainstream outlets as those that 'publish predominantly political content' and are 'popular'. We borrow Twitter-derived domain-level ratings as well as site-visiting statistics from passive metering created in previous work[18,33]. Details on how the two criteria are measured and applied are presented in Methods.

'Trustworthy outlets': These outlets are a particular subset of the set of mainstream media outlets defined above: ones that are annotated as 'trustworthy' in NewsGuard ratings (on the basis of criteria that have to do with journalistic quality and standards of transparency and credibility). We note again that 'regularly publishing false content'—our criterion for deciding the outlets that are considered as fake news outlets in this study—is one of several criteria that determine trustworthiness scores in NewsGuard ratings (details provided in Methods).

'Potentially misleading narratives': These are narratives present in false claims or frequently recurring across content published by fake news sources. These extracted narratives, such as 'vaccines harm people', are not by construction true or false, and cannot be independently judged for truthfulness. They are simply prevalent in misinformation content. However, potentially misleading narratives need not be limited to content present in fake news outlets.

Our approach for automated narrative extraction is detailed in Methods (also summarized in Fig. 3, examples in Table 1). For the sake of being comprehensive, we include results where 'all' structures extracted from fake news content are also used to test our hypothesis, not just recurring ones. However, we define potentially misleading narratives as those narratives that either occur repeatedly in fake news content (where each instance, in trying to mirror mainstream news, contains a detailed story spanning multiple sentences), or the ones that occur in false claims (where each instance is a fact-checked false claim consisting of a short sentence). To enable greater transparency about automatically extracted narrative labels (or string representations), we provide all narratives extracted from false claims (used to test our hypothesis) in Supplementary Text 6 and Supplementary Tables 46–66.

Our initial dataset consists of tweets collected from a panel of over 1.6 million US Twitter users matched to their voter registration records via a commercial US voter file (described in previous work[34]). This panel slightly overrepresents white and female Twitter users, but is otherwise demographically similar to samples of Twitter users collected via high-quality surveys[34,35]. We retain English tweets that share news articles (published by English-language information sources or outlets according to NewsGuard) from 1 May 2018 to 14 November

Table 1 | Examples of extracted narratives

Narrative label	Sentences containing narrative label
baby survive abortion	1. the data on **babies surviving abortions** is compiled on a mandatory basis by only a handful of states.
	2. should a **baby survive** an **abortion**, the health care practitioner shall "exercise the same degree of professional skill, care and diligence to preserve the life and health of the child as a reasonably diligent and conscientious health care practitioner would render to any other child born alive at the same gestational age."
	3. 77% of americans want to protect **babies** who **survive abortions**, nancy pelosi has blocked a vote 80 times
	4. but data from minnesota, the centres for disease control and other sources indicate that **babies** do **survive abortions**, and laws are needed to protect them.
	5. not only are women given the right to sue abortionists for the death of their **babies** that **survive abortions**, those in violation can be criminally charged in federal court for intentionally killing or attempting to kill a human being.
russian hack dnc	1. stone claims if the **russians** didn't **hack** the **dnc**, his case should be dismissed.
	2. the mueller report clearly states that **russians hacked** the **dnc** and gave the hacked emails to wikileaks:
	3. trending: the gateway pundit and american gulag donate $40,000 to persecuted jan. 6 families this christmas on march 8, 2020 and before on june 16, 2019, we presented arguments against the mueller gang's assertion that the **dnc** was **hacked** by **russians**.
	4. the news magazine claimed crowdstrike was correct in assessing that the **dnc** was **hacked** by **russians** in 2016.
	5. when this news hit the media like a bombshell out-of-control left wing judge amy berman jackson tightened the unconstitutional gag order on roger stone to prevent him from discussing the shocking revelation and kravis filed a sur reply with the court falsely claiming that the us government and the mueller investigation had additional evidence to bolster their claim that the **russians hacked** the **dnc**.
bill stop infanticide	1. this is the second time senate democrats have blocked the **bill** to **stop infanticide**.
	2. house democrats again block request to vote on **bill** to **stop infanticide**
	3. nancy pelosi and democrats block **bill** to **stop infanticide** for 75th time, refuse care for babies born alive
	4. democrats block **bill** to **stop infanticide**.
	5. house democrats block request to vote on **bill** to **stop infanticide**

These are the most frequent high-dimensional narrative strings or labels (and sentences containing those labels) in misinformation content that are present in co-shared articles but not present in control articles published by mainstream 'real' news outlets. Terms presented in boldface in sentence text highlight the narrative label's presence in the sentence.

2021. For each tweet, we have the user ID of the author and the news URL being shared in the tweet. We keep URLs that were shared by at least 20 different users (to control for sparsity effects), resulting in ~420,000 different URLs, published by ~2,400 domains and shared by ~450,000 different users.

The methodological approach used in this work is visually illustrated in Fig. 1. Our process for finding co-shared mainstream articles is visualized in Fig. 1b. We construct a weighted graph with individual articles or URLs as nodes. Edges between a fake news URL (nodes shown on the left-hand side of Fig. 1b) and a reliable news URL (nodes shown on the right-hand side of Fig. 1b) are weighted by the number of Twitter users in our dataset that shared 'both' the incident node URLs at least once. We apply a graph-pruning algorithm[36] to assign each edge a co-sharing score on the basis of the likelihood of both incident node URLs getting shared by the same user, controlling for the individual popularity of each of those URLs in isolation. We consider articles published by mainstream outlets that fall into this top 1% set of articles as 'the group of mainstream news articles that are co-shared with fake news'. Figure 2 displays examples of top co-shared articles published by NYT and WAPO, and similarly, Supplementary Fig. 1 displays top co-shared articles published by CNN and Fox News. Stories published by the same set of mainstream outlets that are 'not' in the top 5% of co-sharing likelihood scores form our 'control group'.

We operationalize narratives using methods and an open-source package developed in previous work[30], available at github.com/relatio-nlp/relatio. Figure 3 briefly visualizes the process of extracting the narrative structure through an illustrative example. Extracted narratives are simply represented as strings, or what we call narrative labels or narrative strings ('vaccine cause shingles'; 'vaccine cause people develop shingles'). Two settings are used to extract narratives at two levels of granularity from the text data (following ref. 30): 'low-dimensional narratives', allowing fewer, large-sized clusters or latent entities that result in 'broader' narratives, and 'high-dimensional narratives', allowing a higher number of smaller-sized clusters that result in more specific or 'fine-grained' narratives.

To obtain sets of potentially misleading narratives, we consider two sources of data. The first consists of the aforementioned fake news articles (~24,000 URLs) shared in our Twitter dataset, where we consider the headline and story for each article as separate text instances, since many (at least 6 out of 10) people on social media only read the headline of an article[37]; see also: americanpressinstitute.org/publications/ reports/survey-research/how-americans-get-news/ and washingtonpost.com/news/the-fix/wp/2014/03/19/americans-read-headlines-and-not-much-else/. Headlines and story texts are deduplicated and processed to remove information such as domain name mentions and headline text within the story text. This results in ~30,000 unique texts derived from fake news articles. Potentially misleading narratives in fake news content are defined as recurring narratives: the top 1% in terms of frequency of occurrence (therefore, these narratives occur across several texts in our dataset). We also experiment with all the narratives found in fake news content (Table 2). The second source consists of texts of claims that have been fact checked as false (~24,000 unique texts or claims). For fact-checked false claims, we collect and combine various publicly available datasets (explicated in Methods), and use the claims that are labelled as false (or an equivalent label such as 'not true'). Since these are statements, usually consisting of a single sentence fact checked as false, we consider all narratives extracted from this data as our set of potentially misleading narratives. This results in three libraries of potentially misleading narratives, each containing low- and high-dimensional narratives (Table 2). Table 1 presents examples of potentially misleading narratives.

Methods provides details on the construction of the underlying Twitter panel dataset (including matching between Twitter and US voter data), data collection, sampling of the Twitter dataset, demographic information for the Twitter dataset, process for obtaining news article content, graph-pruning method for obtaining co-shared

Table 2 | Potentially misleading narrative libraries

Source of texts	Potentially misleading narratives library	Description	Total number of narratives (low-dimensional)	Total number of narratives (high-dimensional)
Fake news headlines and stories (30,120 texts)	All fake-news narratives	All narratives found in fake news texts	209,271	246,526
	Recurring fake-news narratives	Subset of above, top 1% in terms of frequency of occurrence	2,544	2,600
False fact-checked claims (21,109 texts)	False-claims narratives	Narratives found in claims that were fact checked as false	10,838	11,384

mainstream articles and constructing the control group, the narrative extraction method and rationales for all the various choices made throughout the method design (such as the choice of date range and thresholds for selecting co-shared articles).

Results

The co-shared and control groups of mainstream news articles and different sets of potentially misleading narratives enable us to test our hypothesis: mainstream news stories that are disproportionately co-shared with fake news are significantly more likely to contain potentially misleading narratives (narratives that are prevalent in misinformation content). Put another way, we test whether potentially misleading narratives occur more frequently in the set of co-shared articles than in the set of control articles published by the same mainstream outlets. Our testing mechanism is visually illustrated in Fig. 1c. We compare the counts or the presence of a set of potentially misleading narratives in co-shared articles with the control set using a one-sided Wilcoxon signed-rank test, a non-parametric significance test for comparing paired or dependent data[38,39] (the procedure for obtaining counts of relevant narratives in particular articles is explained in Methods). Since the 'same' set of narrative labels are being compared for their occurrence in two groups, our scenario fits the paired data testing scenario, and a non-parametric test such as Wilcoxon allows us to proceed without having to assume normality of the underlying distributions. While we use a one-sided test here since our hypothesis is directional (testing that the presence of particular narratives in one group is higher than the other), using a two-sided Wilcoxon signed-rank test replicates all our main results (see Supplementary Text 9 and Supplementary Table 45). Any occurrences of ties and zeros were removed before this test is performed[40], resulting in the final counts or n shown in Table 3.

This process is repeated for all three of our potentially misleading narrative libraries (see Table 2), with the test performed separately for high- and low-dimensional narratives. To control for distributional shifts in the number of articles published by reliable domains, we also perform the same test with articles in both groups limited to those published by trustworthy outlets, as well as articles in both groups limited to those published by liberal outlets alone (the process for obtaining liberal–conservative classification of outlets is described in Methods). This check accounts for the fact that even when the two groups of mainstream articles are published by the same set of domains or outlets, the 'number' of articles published by each domain can constitute different distributions. We show the distribution of per-outlet article counts for the two groups of articles from reliable outlets in Supplementary Text 1 and Supplementary Fig. 2. Since there is a conservative leaning when considering all outlets or even just trustworthy outlets, we repeat our experiments with articles published by liberal outlets alone.

The results across all these settings, shown in Table 3, support our hypothesis: articles from mainstream outlets that are co-shared with fake news contain narratives extracted from misinformation content at a significantly higher rate. This suggests that potentially misleading narratives may allow for a strategic repurposing of mainstream news in fake-news-sharing circles on Twitter, something our data and method can be used to investigate as demonstrated by our case studies (described in the next section). The significance testing procedure is detailed in Methods.

We also estimate the magnitude of the difference in co-shared and control sets across the full set of mainstream news articles, while emphasizing that it is the difference between the two groups that is key to our hypothesis. After combining all recurring narratives in fake news and all narratives in false fact-checked claims as our set of potentially misleading narratives, we count the potentially misleading narratives present in each article. Roughly one (0.94) potentially misleading narrative is present in a co-shared mainstream article on average, whereas for the control set, a potentially misleading narrative is present in only every other article on average (0.58 potentially misleading narrative per article). When dividing by the total number of extracted narratives present in each article, to compare ratios instead of counts, we find that, on average, 2.2% of all narratives in a co-shared article are present in the set of all narratives extracted from misinformation content (recurring across fake news or present in fact-checked false claims), whereas 1.3% of all narratives in a control article are present in the same set of all narratives extracted from misinformation content (recurring across fake news or present in fact-checked false claims). This comparison (average and standard deviation) is also noted in Table 4.

Since we have a large number of texts (more than 50,000 unique texts across fake news articles and headlines, and false fact-checked statements), we are ultimately working with a set of over 27,000 different narrative labels or narratives when checking the number of potentially misleading narratives present in a particular mainstream news article. Although some occurrence of potentially misleading narratives is, therefore, expected in mainstream articles due to factors such as shared language in news reports, our results in Table 3 show that their occurrence is significantly higher in mainstream articles co-shared with fake news articles. Put another way, the value of 0.58 or 1.3% represents the 'noise'—the expected baseline overlap that occurs with mainstream texts due to the methodology of narrative extraction itself—and the '0.94' or 2.2% value for co-shared texts represents the significantly greater 'signal'.

Partisanship as a differing factor between co-shared and control groups

It is possible that these results depend on the co-shared and control articles differing along another dimension other than their propensity to include potentially misleading narratives, namely, partisanship. In Supplementary Text 2, we conduct a series of robustness checks to test for the possibility that the relationships we observe are instead attributable to partisan curation. We find that while right-leaning articles are more likely to contain narratives present in misinformation content, articles in the co-shared group remain significantly more likely to contain those potentially misleading narratives even when controlling for audience partisanship (with and without domain-level fixed effects). Articles in the co-shared group have ~1.24 times the odds of the control group of containing potentially misleading narratives (Supplementary Tables 2 and 3).

Table 3 | Results of our hypothesis tests

Type of outlet	Misinformation narrative library	Dim. of extracted narratives	*n*	One-sided Wilcoxon signed-rank test results
All reliable outlets	All fake news narratives	Low-dim	92,577	Statistic=3,029,034,904.0, ***P=0.0, effect size statistic=∞, 95% confidence interval (CI)=(4.9×10^{-07}, 4.9×10^{-07})
		High-dim	19,845	Statistic=176,302,324.0, ***P=0.0, effect size statistic=∞, 95% CI=(4.9×10^{-07}, 4.9×10^{-07})
	Recurring fake news narratives	Low-dim	1,982	Statistic=1,700,135.0, ***P=8.73×10^{-175}, effect size statistic=0.633, 95% CI=(2.94×10^{-06}, 3.56×10^{-06})
		High-dim	657	Statistic=201,765.0, ***P=3.3×10^{-83}, effect size statistic=0.754, 95% CI=(1.11×10^{-06}, 1.6×10^{-06})
	False-claims narratives	Low-dim	4,413	Statistic=5,132,255.0, **P=0.00095, effect size statistic=0.05, 95% CI=(4.9×10^{-07}, 4.9×10^{-07})
		High-dim	318	Statistic=32,746.0, ***P=2.97×10^{-06}, effect size statistic=0.262, 95% CI=(4.9×10^{-07}, 4.9×10^{-07})
Trustworthy outlets	All fake news narratives	Low-dim	86,871	Statistic=2,533,598,901.0, ***P=0.0, effect size statistic=∞, 95% CI=(5.62×10^{-07}, 5.62×10^{-07})
		High-dim	15,912	Statistic=109,142,619.0, ***P=0.0, effect size statistic=∞, 95% CI=(5.62×10^{-07}, 5.62×10^{-07})
	Recurring fake news narratives	Low-dim	1,924	Statistic=1,587,705.0, ***P=1.09×10^{-162}, effect size statistic=0.62, 95% CI=(2.76×10^{-06}, 3.37×10^{-06})
		High-dim	587	Statistic=159,336.0, ***P=3.09×10^{-71}, effect size statistic=0.737, 95% CI=(1.12×10^{-06}, 1.69×10^{-06})
	False-claims narratives	Low-dim	4,281	Statistic=4,482,215.0, P=0.89, effect size statistic=0.002, 95% CI=(4.68×10^{-07}, 5.62×10^{-07})
		High-dim	306	Statistic=29,136.0, ***P=0.00012, effect size statistic=0.22, 95% CI=(5.62×10^{-07}, 5.62×10^{-07})
Liberal outlets	All fake news narratives	Low-dim	71,654	Statistic=1,831,440,473.0, ***P=0.0, effect size statistic=∞, 95% CI=(9.2×10^{-07}, 9.2×10^{-07})
		High-dim	8,235	Statistic=29,733,384.0, ***P=0.0, effect size statistic=∞, 95% CI=(9.2×10^{-07}, 9.2×10^{-07})
	Recurring fake news narratives	Low-dim	1,856	Statistic=1,439,940.0, ***P=9.26×10^{-139}, effect size statistic=0.582, 95% CI=(2.76×10^{-06}, 2.83×10^{-06})
		High-dim	396	Statistic=70,018.0, ***P= 7.59×10^{-42}, effect size statistic=0.681, 95% CI=(1.84×10^{-06}, 1.85×10^{-06})
	False-claims narratives	Low-dim	3,956	Statistic=4,442,138.0, ***P=8.41×10^{-14}, effect size statistic=0.119, 95% CI=(-8.998×10^{-07}, -8.65×10^{-07})
		High-dim	263	Statistic=24,814.0, ***P=6.37×10^{-10}, effect size statistic=0.381, 95% CI=(9.2×10^{-07}, 9.2×10^{-07})

These results show that across various categories of reliable mainstream news outlets and sets of potentially misleading narratives, potentially misleading narratives are present at a higher rate in co-shared articles than in the control group. *P<0.05, **P<0.005, ***P<0.0005. Dim. refers to the setting or type of dimensionality used when extracting narratives: low- or high-dimensional.

Case studies

Our quantitative results show that narratives are significantly associated with co-sharing of certain mainstream articles in the 'aggregate'. In addition, our data and methods can help contextualize the relationship between information and narratives, and how particular mainstream articles co-shared with fake news can be used or repurposed to spread potentially misleading narratives. To this end, we conduct two case studies, covering two different topic areas: vaccine-related misinformation and voter-fraud-related misinformation, both prominent on Twitter and online social media in 2020[26–28]. These case studies highlight how our work can contribute to the understanding of specific narratives in the media landscape, as well as enable journalists and scholars to help investigate potential cases of synergistic support for potentially misleading narratives in mainstream news.

Case Study 1 (Vaccination)

In our manual examination of narrative labels (and narratives) present in co-shared articles published by liberal mainstream outlets, we find multiple labels of the form 'vaccinated people spread disease', 'vaccinated people spread virus' and 'vaccinated people spread delta variant'. We present details of our manual filtering approach in Supplementary Text 3, and all such narrative labels obtained along with sentences containing them in Supplementary Tables 6–8. Here we examine the nine mainstream articles that are (1) published by 'liberal' mainstream outlets; (2) co-shared with fake news; and (3) contain the aforementioned identified type of narrative. The article URLs, their headlines, relevant story snippets and relevant narrative labels they contain are provided in Supplementary Tables 9–11. These nine articles can be grouped qualitatively into three types in terms of their propensity to be used for misleading purposes:

- Type 1: Clickbait headlines. We find that three mainstream articles contain at least a mild form of clickbait or misleading headlines, via trivialization or simplification that could provide easy support for misleading anti-vaccination narratives and beliefs. Headlines carry particular significance in their much wider reach and readership[37,41] as well as their influence on how the article, if read, is interpreted[42]. When written for the primary purpose of generating clicks rather than to inform, headlines can omit context in ways that sensationalize by presenting the news item

Table 4 | Quantifying presence of potentially misleading narratives in the two groups of mainstream articles

	% of narratives in an article that are present in the potentially misleading narrative set (average (s.d.))
Co-shared mainstream news group	2.2% (4.2)
Control mainstream news group	1.3% (3.3)

as "more interesting, extraordinary and relevant than might be the case"[43], and such clickbait headlines can be actively misleading[44]. A recent study suggests that the pragmatic implications of news headlines influence the readers' reactions, and reasoning about these implications of headlines is important to combat misinformation[45]. For example, the headline "Top health expert says vaccinated people are spreading delta variant" (published by The Hill; first row in Supplementary Table 9) can be used to ascribe a causal or unique role to vaccinated people in the spread of the variant as opposed to unvaccinated people (who are not implicated via omission). The article itself is careful not to suggest this, noting that "public health experts are encouraging further vaccinations to help curb transmission. Some experts, *however*, warn that vaccinated individuals *may still* be capable of contracting and transmitting COVID-19" (emphasis added). The addition of uncertainty is in direct contrast to the headline. Further, this snippet clarifies that rather than having a unique or causal role in transmissions, vaccines do not completely eliminate transmissions. This case is related to that of incongruent headlines[46], when statement(s) in the headline are not supported in the article's content, thereby misleading readers. This case also features an example of what linguists would call a quantity implicature[47], in which an ambiguous statement (for example 'Q: Who came to the party? A: Amy and Bonnie') is interpreted to exclude other possible alternatives (implying that only Amy and Bonnie came to the party, although the answer is true if additional people also came). Here, the phrasing of the headline permits and arguably encourages restrictive interpretations in which vaccinated people are spreading the delta variant at equal or higher rates than unvaccinated people, supporting a narrative in which vaccines are ineffective or actively harmful, despite the article itself making clear this is not the case. The use of the headline to support anti-vaccine worldviews is further illustrated in the text of the tweets sharing these articles (Supplementary Table 13), such as 'Vaccines that target the spike protein are causing it to mutate. Virology 101…' and '…THAT IS LITERALLY THE HEADLINE. I wonder if President Biden will consider me a "murderer" for sharing'.

- Type 2: Older news articles in new contexts. These are articles published well before the period we analyse (May 2018 to November 2021) that report on particular vaccine failures in the past, reintroduced by Twitter users to make claims about contemporary vaccines. As shown in Supplementary Table 10, we find two such articles, one from mid-2017 by NPR (headline: "Mutant Strains Of Polio Vaccine Now Cause More Paralysis Than Wild Polio") and another from 2015 by PBS (headline: "This chicken vaccine makes its virus more dangerous"). Sharing older news articles in a new context can be used to mislead or as a disinformation tactic; news organizations have taken steps to highlight their publication date[48]. In 2020, Facebook rolled out a feature to notify users when they are sharing articles more than 90 days old (see dailymail.co.uk/sciencetech/article-8460273/Facebook-tell-users-share-old-news-article.html and forbes.com/sites/alisondurkee/2020/06/26/new-facebook-policy-warns-users-before-they-share-old- articles/?sh=3b0eb9fb3f34). Tweets by users promoting vaccine misinformation sharing these articles (Supplementary Table 13) can further help understand such repurposing of previous, legitimate reporting. For example, the aforementioned polio vaccine article by NPR was shared with comments such as "None for me, thanks. I'll take my polio natural." and "2 year old article is still relevant., Nope. Vaccines have been causing VARIANTS for a LONG time. All you folks that got VACCINATED are the ones who have "caused this". #TheVaccinesAreTheVariants."
- Type 3: Other. These are the remaining articles that do not seem to be obviously usable for misleading purposes.

We show a sample of the vaccine-related fake news articles co-shared with the aforementioned mainstream articles in Supplementary Table 12. This provides exposure to the interactions between fake news and co-shared mainstream articles. A random sample of tweets from users promoting vaccine-related fake news who also shared these articles from liberal-leaning mainstream outlets are provided in Supplementary Table 13, while 'all' the unique tweet text from users who shared vaccine-related fake news and both types of mainstream articles are also presented in Supplementary Text 3. These tweets illustrate 'how' these 'real' news articles might be getting used to support potentially misleading narratives.

Finally, we take a closer look at the three mainstream articles (Supplementary Table 9) with clickbait headlines and quantify their impact in terms of their disproportionate circulation among users promoting vaccine misinformation (by having a history of sharing vaccine-related fake news articles). We find that these particular articles by the three mainstream outlets (The Hill, Business Insider and WPO) are some of their most popular vaccine-related articles among these users, relative to all the vaccine reporting done by those outlets (based on the URLs shared in our Twitter dataset between May 2018 and November 2021). For example, out of all the 444 vaccine-related articles published by WAPO in our final dataset, the particular co-shared article with the headline "CDC study shows three-fourths of people infected in Massachusetts coronavirus outbreak were vaccinated but few required hospitalization" was the 2nd highest in terms of shares among the subset of Twitter users known to share vaccine-related fake news (washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated). We also find that these articles were disproportionately shared by users who also shared vaccine-related fake news, relative to a comparison group of users who shared reliable vaccine information.

We provide all relevant sets of articles and tweet text along with details of various procedures involved in this case study in Supplementary Text 3, including the process of obtaining the vaccine-related fake news articles co-shared with mainstream articles of interest and significance tests that highlight the disproportionate uptake of these mainstream articles with clickbait headlines among users who share vaccine-related fake news.

Case Study 2 (Voter fraud allegations in the 2020 US elections)

One of the most widespread mis- or disinformation campaigns of 2020 in the United States was allegations of voter fraud in the 2020 US election, which can negatively impact people's trust and confidence in the election process[27,28]. Unsurprisingly, we find many narrative labels capturing claims of voter fraud, especially focusing on the use of mail-in ballots since mail-in-based voting was significantly expanded in the 2020 elections in the wake of the COVID-19 pandemic and social distancing recommendations. Similar to case study 1 above, we conduct an examination of co-shared mainstream articles published by liberal outlets containing narratives that could capture voter fraud claims and analyse the possibility of some mainstream news articles providing potential ammunition to users who have a history of sharing fake news.

We compile relevant extracted narratives by first filtering all narrative labels containing 'fraud' and 'mail', 'faulty' and 'ballot' and other keywords that emerge upon manual examination such as 'behalf'; we also filter out irrelevant narratives that directly suggest a focus on 'allege' and 'speculate' instead of asserting a claim. We manually select similar relevant and coherent narratives together under an umbrella for this case study, allowing for a lack of precision to try to capture maximal cases of such narratives and potential articles to examine qualitatively. We provide details of our process, all 150 narratives that emerged via keyword search, and all 49 relevant narratives that emerge upon manual examination, along with sentences containing those selected narratives in Supplementary Text 4. These narratives are then used to find relevant mainstream co-shared articles.

Specifically, we manually examine all mainstream articles that are (1) published by 'liberal' mainstream outlets; (2) co-shared with fake news; and (3) contain at least one of the selected narratives that pertain to voter fraud. For all these mainstream articles, the article URLs, their headlines, relevant story snippets and relevant narrative labels they contain are provided in Supplementary Tables 19–22. After examining these articles, we find 11 instances of articles that could potentially be used to support a narrative or worldview of widespread voter fraud. These 11 articles can be grouped qualitatively into three types, in terms of how they might be useful for misleading purposes:

- Type 1: Mainstream articles easy to use for promoting mail-in-vote-based fraud conspiracy theory. These articles are interesting cases where the content, including the headline, provides easy use or re-use for promoting the voter fraud narrative in 2020 in the United States: (1) nytimes.com/2012/10/07/us/politics/as-more-vote-by-mail-faulty-ballots-could-impact-elections.html (with the headline "As More Vote by Mail, Faulty Ballots Could Impact Elections"); (2) washingtonpost.com/history/2020/08/22/mail-in-voting-civil-war-election- conspiracy-lincoln (with the headline "Mail-in ballots were part of a plot to deny Lincoln reelection in 1864"); (3) thehill.com/opinion/campaign/506331-vote-by-mail-would-create-chaos-and-distrust-in-november (with the headline "Vote-by-mail would create chaos and distrust in November"). The report from The New York TIMES was published in 2012, but given its suggestion of vote-by-mail increasing chances of faulty ballots and having an impact on elections, it provides a perfect mainstream story to support the voter fraud conspiracy in 2020. This is also an example of temporal repurposing or using an article from a different time and context (as discussed about 'type 2' articles in case study 1), to promote a certain narrative in a different time and context. The article's story content clarifies that the number and impact of mail-in ballots is not as clear cut, but given the strong push in 2020 toward a belief in fraud and questioning the election outcome even before it occurred, even just the headline lends itself to the conspiracy and can be 'easily' used towards this end (indeed, as we establish below, since this article from The New York Times is disproportionally popular among voter-fraud or mail-in ballot-related misinformation sharers on Twitter in our dataset, compared with all reporting on voter fraud from The New York Times in our dataset). We can see strategic use of this article, including emphasis on the domain name (The New York Times) to take advantage of the outlet's reputation, in tweets sharing this article by users who have shared voter-fraud or mail-in-ballot-related fake news; for example, the tweet "Even The NY Times agreed that vote by mail ballots could impact the 2012 election. So why can't they impact the election in 2020? Well." and the tweet "FLASHBACK NEW YORK TIMES (2012): As More Vote by Mail, Faulty Ballots Could Impact Elections hat tip:". The Washington Post article similarly can be easily and directly used to promote voter fraud conspiracy theories, especially those centred on mail-in ballots. The headline itself: "Mail-in ballots were part of a plot to deny Lincoln reelection in 1864" clearly states mail-in ballots were used to try to influence an election outcome. Perhaps alarmingly, this article was published in August 2020, when misinformation about mail-in-ballot-based fraud and its exaggerated impacts on election outcomes had plenty of momentum, especially on social media. Furthermore, this article and headline focuses on a deliberate, intentional plot to impact an election, and intentional fraud and 'stealing the election' is at the heart of the conspiracy in 2020. While the subject of this story is an election that occurred in 1864, long before 2020, it provides an instance of a story from a mainstream outlet that can be easily used to prop up misinformation around electoral fraud in 2020. In fact, given the timing of this article and a focus on the intent to impact an election, this might be the most useful mainstream article for supporting the conspiracy that we have uncovered in our data. Finally, The Hill article is an opinion piece, making a negative case for vote-by-mail, conceding the lack of trust and 'chaos' it would create. It also claims mail-in voting is a cause of fraud, surfacing select research while ignoring other research. Since a mainstream outlet is making the case 'against' vote-by-mail, it can be directly and easily used by people interested in using expanded mail-in voting to suggest electoral fraud.
- Type 2: Mainstream articles reporting on actual cases of voter fraud, or allegations with a concrete basis. The thrust behind the voter fraud conspiracy theory for the 2020 election is not that individual-level fraud or mistakes never occur, but that such fraud has occurred or will occur in a 'widespread' manner and at a high enough rate as to impact the election, which is the false claim or belief. Although the claim asserts a large magnitude, people making this false claim can use individual cases of actual fraud, or concretely alleged fraud, as 'proof' of the alleged widespread nature of the phenomenon. To this end, mainstream reports on individual cases of fraud or alleged fraud are useful for misinformation spreaders as something concrete and believable to point to, reported by reputable mainstream outlets. We find six such articles: wsaz.com/2020/07/10/mail-carrier-in-west- virginia-pleads-guilty-to-attempted-election-fraud

  (1) (with the headline "Mail carrier in West Virginia pleads guilty to attempted election fraud"); kron4.com/news/california/northern-california-woman-faces-felony-voter-fraud-charges-for-allegedly-voting-twice/1575491353
  (2) (with the headline "Northern California woman faces felony voter fraud charges for allegedly voting twice"); nbcnews.com/news/us-news/republican-official-ohio-faces-charge-voting-twice-november-election-n1271985
  (3) (with the headline "Republican official in Ohio faces charge for voting twice in November election"); politico.com/story/2018/12/04/north-carolina-elections-fraud-allegations-mark-harris- campaign-house-1045355
  (4) (with the headline "GOP hit with election fraud claims after using issue as rallying cry"); businessinsider.com/voter-election-fraud-pennsylvania-charge-dead-mom-vote-trump-2020-12
  (5) (with the headline "Officials finally found a case of a dead person voting, accusing a Republican of pretending to be his dead mom to vote for Trump"); thehill.com/homenews/531242-pennsylvania-trump- supporter-charged-with-voter-fraud
  (6) (with the headline "Pennsylvania Trump supporter charged with voter fraud").

  Observing tweets by users who have a history of sharing voter fraud-related fake news articles, we see the use of these individual cases to push for the widespread nature or push back against the refutation of conspiracy theory by mocking the idea that fraud cannot happen; for example, a user with a history of sharing voter fraud-related fake

news shared the WSAZ News article (on a mail carrier in West Virginia pleading guilty to attempted election fraud) with the tweet: "But but but this would never really happen!!! Mail carrier in West Virginia pleads guilty to attempted election fraud for attempting to alter mail-in ballot request forms."

- Type 3: Mainstream articles containing mail-in-ballot fraud claims from actors, without any pushback. We also find two mainstream news articles that contain the false voter fraud claim (via some actor making the claim and getting quoted on that) without any pushback or fact check for the claim. In this way, the false claim is allowed to stand without providing any correction in a mainstream news report, which can potentially legitimize the false claim or be used for such legitimization by actors who believe the claim and want to spread it. The two such articles we find are: thehill.com/homenews/house/492057-mccarthy-slams-democrats-on-funding-for-mail-in-balloting (with the headline "McCarthy slams Democrats on funding for mail-in balloting") and nbcnews.com/politics/meet-the-press/two-thirds- voters-back-vote-mail-november-2020-n1187976 (with the headline "Two-thirds of voters back vote-by-mail in November 2020").

We show a sample of the voter-fraud or mail-in-ballot-related fake news articles co-shared with the aforementioned mainstream articles in Supplementary Tables 23 and 24. This provides further examples of the interactions between fake news and co-shared mainstream articles. A random sample of tweets from users promoting voter-fraud or mail-in-ballot-related fake news who also shared these (type 1 or type 2) articles from liberal-leaning mainstream outlets are provided in Supplementary Table 25. These tweets illustrate 'how' these 'real' news articles might be getting used to support potentially misleading narratives.

Finally, similar to case study 1, we take a closer look at the three type-1 mainstream articles (Supplementary Table 19) that can be very easily and directly used to promote the voter-fraud (particularly, mail-in-vote-based fraud) conspiracy theory about the 2020 US election, and quantify their impact in terms of their disproportionate circulation among users promoting voter-fraud or mail-in-ballot-related misinformation. We find that these particular articles by the three mainstream outlets (The New York Times, The Washington Post and The Hill) are the respective outlet's most popular voter-fraud or mail-in-ballot-related articles among these users, relative to all the voter-fraud or mail-in-ballot reporting done by those outlets (based on the URLs shared in our Twitter dataset between May 2018 and November 2021). For example, out of all the 99 voter-fraud or mail-in-ballot-related articles published by The Washington Post in our final dataset, the particular co-shared article with the headline "Mail-in ballots were part of a plot to deny Lincoln reelection in 1864" was ranked the highest in terms of shares among the subset of Twitter users known to share voter-fraud or mail-in-ballot-related fake news. We also find that these articles were disproportionately shared by users who also shared voter-fraud or mail-in-ballot-related fake news, relative to a control group of users who shared reliable voter-fraud or mail-in-ballot-related information.

We provide all relevant sets of articles and tweet text along with details of various procedures involved in this case study in Supplementary Text 4, including the process of obtaining the voter-fraud-related fake news articles co-shared with mainstream articles of interest and significance tests that highlight the disproportionate uptake of clickbait mainstream articles among users who share voter-fraud-related fake news.

Discussion

Consistent with our hypothesis, we find that mainstream articles that circulate among fake-news-sharing users on Twitter are significantly more likely to contain narratives that are prevalent in misinformation content, compared with mainstream articles from the same reliable outlets that are not co-shared with fake news. This effect is not fully attributable to partisan curation—our findings are not entirely dependent on potentially misleading narratives carrying a left- or right-leaning slant. Our finding suggests that users strategically repurpose mainstream news to develop and spread potentially misleading narratives on social media. Indeed, it is likely that users promoting misleading narratives find mainstream sources particularly attractive when they publish information that can be repurposed to fit those narratives 'precisely because of' these sources' credibility. This is consistent with related instances of partisan information being seen as more credible when it comes from unexpected sources[49], and suggests a general framework in which users share information that they perceive to be 'useful' for advancing their broader worldviews.

Analyses based on our dataset show that the reach and potential audience for mainstream news articles co-shared with those fake news articles is 'more than twice' that of fake news articles. Specifically, on average, ~121 users share the co-shared mainstream articles as compared with ~58 who share a given fake news article. Moreover, those who exclusively share fake news average ~23 followers, while those who exclusively share co-shared mainstream articles averaged ~45 followers (details are provided in Supplementary Text 7; the difference between the two groups is statistically significant). This nearly 2:1 ratio is also observed when comparing the median numbers for (co-shared) mainstream and fake news article sets. This further underlines the need to study mainstream news sources and various articles published by reliable outlets, along with their interaction with the relatively niche presence of, and potential exposure to, fake news sources and articles. Potentially misleading narratives, by cross-cutting the type of outlet (fake news or mainstream), have a much larger platform than that of fake news alone.

To the extent to which our finding about the role of mainstream news in online misinformation networks on social media is true, it has important implications for responsible journalistic practice: when vetting news stories and their framing, especially for the headlines, it is important to consider not just the raw information of the story itself but also the broader claims that the information could be used to support. Technological methods such as the ones we have developed and used here may help go beyond checking a story's factuality and accuracy and contribute to minimizing the risk of a story being repurposed to mislead. Journalists committed to the rigour of their reporting and how their content might get used or misused can adjust their framing accordingly before publishing. This can help reduce the chances of mainstream news articles potentially legitimizing worldviews based on false or unsubstantiated claims.

For scientific research on online misinformation, our work highlights the importance of considering the broader context in which a specific piece of information is being circulated, rather than solely relying on classifying information (to say nothing of sources) as strictly reliable or unreliable. This is connected to the idea of 'deep stories', the larger underlying stories (or narratives) tying together various atomized daily stories occurring in people's lives[50], which can often be 'the' connective force behind people's beliefs regardless of plausibility and has been tied to fake news and disinformation in the United States[23,51]. It is also important to examine the choices being made in mainstream news, such as the inclusion of existing misleading narratives. Indeed, the focus on the purely informational component of misinformation is insufficient to account for why misperceptions are widespread and persistent, especially given the small market share of categorically 'fake' news outlets. Instead, a better framework for understanding the apparent problems with the online information ecosystem must account for user preferences and behaviour, including their preferences for information for reasons other than its truth value (such as its usefulness for protecting worldviews and advancing interests)[52].

Very few users share articles from known fake news outlets[4,53]. One possible reason is that users perceive reputational harm from sharing information from unreliable sources[53]. If the same belief can be

supported with a mainstream source, rather than an unreliable source, it is likely to be preferred to avoid this reputational cost. Articles by mainstream outlets that are co-shared with fake news (such as The Washington Post article highlighted in Fig. 1 originally titled "Vaccinated people now make up a majority of covid deaths") might become a part of the rationalization process for misperceptions and misleading narratives or beliefs. Our case studies suggest as much, with tweets such as "Well look what the CDC is saying today. It's the vaccinated spreading covid." sharing the article businessinsider.com/cdc-fully-vaccinated-people-can-spread-delta-variant-2021-7 with the headline "CDC says fully vaccinated people spread the Delta variant and should wear masks: 'This new science is worrisome'." The same misleading narrative can now be rationalized using a combination of different information sources (fake news, mainstream news, cross-partisan outlets).

Existing narratives present in misinformation content being also present in the mainstream media landscape also hints at the potential agenda-setting power of fake news. Previous work has uncovered a temporal alignment in news volume and bursts of coverage between fake (or unreliable) and reliable news ecosystems, which suggests a competition for setting the agenda of public discourse[54]. Fake news content can both respond to, and influence, the agenda in more mainstream partisan media[55], and mainstream news outlets are sometimes responsible for disseminating and popularizing particular fake news stories[56]. Another form of interaction between fake and mainstream media ecosystems is fake news outlets potentially capitalizing on the sensationalism 'already' present in the mainstream media environment[57]. As also highlighted by our study, this underscores the responsibility and need for structural action in the mainstream journalism landscape itself rather than confining the problem of misinformation to a certain set of identified 'bad actors' in the form of specific fake news outlets.

Our findings also complement recent work that more specifically examines false and misleading narratives regarding the COVID-19 vaccine within communities on social media[21]. Not only are the identified narratives similar, but the authors in that work noted the "distortion of facts due to misleading rather than outright false narratives" such as exaggeration of side effects and recontextualization. This mirrors our findings in terms of observed extracted narratives, the co-shared fake and mainstream news content, and the indication of strategic repurposing of mainstream news content on social media (which can help support efforts to recontextualize legitimate reporting). The latter indication of repurposing and possibly taking advantage of how headlines and mainstream story content are framed is backed by the findings of our case studies. These findings also relate to recent scholarly understanding that the challenge of misinformation on social media is less about 'bad facts' and more that people 'are more often misled not by false evidence but by misinterpretations and mischaracterizations—dynamics of a collective sensemaking process gone awry'[58].

Limitations and future work

For our quantitative results in this study, we examine the content of articles, but for social media posts, we only considered the articles shared and not the text generated by the user while sharing the article. Our findings and our case studies suggest that future work should also look at the 'content of tweets' sharing news articles, robustly identify misinformation sharers, and study 'how' mainstream news articles get repurposed to support pre-existing misperceptions via the potential choices made in tweet texts (while maintaining an understanding of the larger media ecosystem that enables such repurposing). Examining post content can also help verify what our findings indicate: mainstream outlets might be used to grant an air of legitimacy to certain narratives. Such an examination can help analyse the instances where mainstream news content itself misleads 'directly' and instances where mainstream news tends to 'enable' misleading usage by social media users. Furthermore, while our approach in this work focuses only on external links shared on Twitter, false and misleading claims can and do often circulate in social media posts without external references or links; however, our approach is centred on the understanding that it is the credibility of the external reference that misinformation sharers capitalize on.

Studying the content of the social media post sharing an article can also help assess whether the article is being shared primarily to criticize the article, the outlet, or a particular type of media, for example, by using ironical commentary such as 'look at what the New York Times is saying now'. To help assess the possibility of this 'share-to-criticize' pattern in our dataset, we conducted a manual annotation analysis of 100 randomly selected tweets that share an article published by a trustworthy source, where the tweet is authored by a Twitter user with a history of sharing at least five distinct fake news articles. Three independent annotators labelled each tweet as to whether it was criticizing the article it was sharing (or if such a determination cannot be made without additional context). A majority of the annotators labelled the tweet as sharing the article to critique or dispute it (or the outlet or the type of media) in only 3 cases. This suggests that the 'share-to-criticize' behaviour is very limited in our dataset and unlikely to have influenced our results (more details on the human annotation task and analysis are provided in Supplementary Text 10). However, future work should investigate the broader concern in which misinformation sharers share mainstream news articles on social media in a critical or ironic manner.

Another subject for future analysis would be to examine where and how misleading narratives occur in mainstream news stories. One possibility is that these narratives do not occur in the content written by the journalist or news reporter themselves, but in 'direct quotations': content attributed to a different actor or source replicated and used verbatim in a text. Such direct quotations are explicitly marked using quotation markers (such as "…") in texts[59]. We removed instances of direct quotations from co-shared and control articles in a preliminary analysis (using patterns that err conservatively toward removing rather than keeping potential directly quoted text), and found that our hypothesis still holds when working with this modified content. Therefore, the significantly higher presence of these potentially misleading narratives in co-shared articles does not seem to be dependent on the practice of direct quotations in media alone. Overall, our effect holds for these newly processed texts (with direct quotations removed) in 16/18 significance testing instances (where each testing instance corresponds to one row in Table 3). We report our preliminary analysis including the procedure to remove direct quotations and detailed results in Supplementary Text 5.

Nonetheless, future work should examine whether the practice of direct quotations (and the possible presence of misleading narratives in those quotes) necessarily abdicates the responsibility of journalists. Direct quotations still constitute a journalistic choice; indeed, one previous work argues that journalistic quotation (direct or indirect) is a 'highly interpretive compositional activity', making it a subjective part of creating a news story[60]. Direct quotations can constitute a 'shift' in responsibility to the source as well as a shift in the interpretation or attitude towards the quoted content as not a part of the factual reporting of events[61]. However, even if this is the case, it is possible that in the context of strategic repurposing of mainstream news, misinformation conveyors, as well as their audience, might be able to ignore this shift. Future work can also examine the effect of indirect quotation, where quotation markers are not used and instead, quotes are presented using reporting verbs such as 'said' ('they *said* that…*reported content* …'). Such an investigation can help further understand the distribution of where misleading narratives occur in mainstream media reporting, when citing a source verbatim or in original writing by the author of the news article.

Our work uses one particular way to categorize outlets as fake news—outlets annotated as 'regularly publishing false content'—aiming for precision. Future work can experiment with a broader set of

unreliable content to study the interactions between reliable and unreliable content.

By scoring individual news articles for their co-sharing likelihood, our work can also help inspire alternatives to the dominant research paradigm that uses domain-level reliability ratings and tends to treat all individual articles published by the same outlet in the same manner. We note that we also treat articles published by 'fake' news outlets in that manner, but we consider each individual story published by 'mainstream' outlets to identify cases of misinformation adjacency. We also encourage future work to further refine the set of potentially misleading narratives obtained and used in this study to help answer specific research questions and conduct additional investigations of specific misleading narratives. In Supplementary Tables 46–66, we provide all the high-dimensional narratives automatically extracted from the texts of fact-checked false claims and used in the comparison of their presence in co-shared and control mainstream articles (for the results reported in Table 3).

While we quantify the relative presence of potentially misleading narratives in the co-shared and control groups of mainstream articles, our focus in this work is to compare and test the 'difference' among the two groups, and not quantify the presence of possible misinformation in various mainstream media articles. The specific values for the presence of potentially misleading narratives in mainstream articles rely on the method itself, which in turn relies on keywords and extracts many narrative structures from any piece of text. To quantify the presence of potentially misleading narratives in mainstream media more accurately, future work would need to propose a more precise method for that particular quantitative estimation, which is out-of-scope for our work. That future work can certainly build on our work, especially our research design, to identify articles of interest in the quest for quantifying the amount of misleading framing in misinformation articles.

The narrative extraction tool we use first splits the corpus into sentences and then works at the sentence level. This limits the context to just the sentence, which can result in some artefacts. For example, while negations within a sentence are captured, if someone presents a claim in an article only to then negate it later on in a different sentence, this will not be captured by the method.

Finally, we note that contending with other explanations can clarify the relative role of potentially misleading narratives and also serve as robustness checks for our work. Here we accounted for the potential relevance of partisan curation of news articles (Supplementary Text 2). However, other possible distinctions between the co-shared set of articles and our control group can be the relative presence of certain entities, or even interest in specific topics that may dominate the fake news world and co-shared mainstream articles. Conducting our hypothesis tests within subsets of articles based on the topic or theme in the news content, and also subsets defined by the presence of prominent entities, constitute interesting directions for future research enabled by our processed dataset.

Methods

The Twitter dataset used in this study was previously approved for research[34,35] and collection by the Institutional Review Board (IRB) at Northeastern University (17-12-13).

Construction of the underlying Twitter dataset and data collection

Our initial dataset consists of tweets collected for a panel of over 1.6 million US Twitter users that are matched with their voter registration records, obtained via a commercial voter file provided by TargetSmart (described in previous work[34]). This panel slightly overrepresents white and female Twitter users, but is otherwise demographically similar to samples of Twitter users collected via high-quality surveys[34,35]. This panel is appropriate for our study as we are chiefly concerned with (mis)information shared on social media, for which the sharing behaviour of users is necessary, and the ability to verify that these users are real people is extremely useful. The sample was constructed by identifying users active in the Twitter Decahose data (a random 10% sample of all of Twitter that was available to academics via the Twitter API before 2022) with a unique first name/last name/location that also appeared in a commercial voter file. In other words, Twitter users were exact matched to a commercial voter file by first name, last name and location. All matches were retained in the panel (that is, we did not randomly sample). After the user panel was constructed, users' tweets were collected at regular intervals using the Twitter API. Further details about construction and validation of this dataset are provided in previous work[34,35].

We retained English tweets that shared news articles (published by English-language information sources or outlets based on NewsGuard) from 1 May to 14 November 2021. This resulted in an initial dataset containing ~14.6 million unique URLs published by ~4,000 domains shared by ~600,000 different users. For each tweet, we had the user ID of the author and the news URL being shared in the tweet. We retained URLs that were shared by at least 20 different users (to control for sparsity effects; we did not conduct further sampling as our study is large-scale observational), resulting in ~420,000 different URLs published by ~2,400 domains and shared by ~450,000 different users. Compared with the full panel of ~1.6 million US Twitter users, our subset of ~450,000 users slightly overrepresents white and female Twitter users, and overrepresents users registered as Democrats. Our subsample also slightly underrepresents Twitter users in the 18–29 age bucket and, in almost equal proportion, slightly overrepresents users in the 50–64 age bucket.

Our choice of date range was influenced by two internal criteria:

(1) Use of the latest Twitter panel data at the time of the beginning of experimentation and research design (which began in mid 2021 and was revisited at the end of 2021 when our methodology took concrete shape).
(2) Computer-related constraints, such as memory, for loading and working with large datasets.

This is how our date range ended up being May 2018 to November 2021: we tried to go back in time from November 2021 as much as computationally feasible. The reasoning behind trying to work with the latest data at the time of initial experimentation was to retain the latest narratives and shared news in a fast-evolving political climate. Our date range encompasses the whole of 2020, which was a unique year due to the onset of a once-in-a-century pandemic of COVID-19 and also the occurrence of a US presidential election, seeing an active and charged political climate throughout. As such, it is unsurprising that 2020 news stories seem to feature prominently in our work.

As noted above, the data for this study came from a pre-existing panel of Twitter users matched to a commercial voter file. The commercial US voter file was obtained from TargetSmart in October 2017. Twitter API v.1 was used to collect Twitter data. The text data of news articles were collected using the Newspaper3k API (v.0.2.8) in Python as a Python library (newspaper.readthedocs.io/en/latest/). NewsGuard data (another commercial resource) provided classifications of domains as detailed below.

Use of NewsGuard annotations

NewsGuard is an independent organization that involves a team of journalists evaluating the quality of online information sources. NewsGuard uses web tracking data along with ratings by professional news editors and journalists (a trained team of experts) to produce the final sets of annotations and scores at the domain level. The panel of editors and journalists offer a trustworthiness score between 0 and 100 which is based on several different criteria, and annotations for those criteria are also included. These criteria have to do with journalistic quality and standards of transparency and credibility. As such, the criteria include

publishing of false information, fact-checking standards of the news institution, disclosing advertising agencies and all sources of funding, being transparent about ownership, level of editorial oversight and so on. Having annotations for these different criteria and not just a score or a 'blocklist' of domains is useful in making finer-grained assessments and allows for nuance. For example, less-reliable domains can be distinguished on the basis of the potential reason for a deficiency in quality: not disclosing ownership details is qualitatively a different reason to not trust a website than the regular publishing of false information. Detailed descriptions of all journalistic criteria involved in the ratings are available at https://www.newsguardtech.com/ratings/rating-process-criteria/.

NewsGuard's domain-level judgments show a high overlap with other lists with publisher ratings[62]; specifically, recent work comparing six sets of news domains' quality ratings (including NewsGuard ratings) found that "they generally correlated highly with one another"[63] (see also refs. 64,65). NewsGuard's ratings and annotations are commonly used in computational misinformation research[64–66], albeit in different ways, such as using the provided 'trustworthiness' scores to divide information sources into groups[64,65]. NewsGuard has sought to add news sources over time, and its popularity in misinformation research warrants an independent investigation of its content, processes and use. This is the goal of a recent article[66], with the manuscript made publicly available as a resource examining NewsGuard annotation data in detail (osf.io/preprints/psyarxiv/v6e4b).

In this work, we aimed for precision by considering 'regularly publishing false content' as our criterion for demarcating fake news outlets, out of all the factors that determine trustworthiness in NewsGuard ratings. This helped highlight the drawbacks of conceptualizing problems with the digital information ecosystem as that of 'false information' or fake news content, a framework that often foregoes closer examination of information coming from sources deemed reliable. All other outlets were deemed 'reliable'. We used this specific annotation for domains in NewsGuard data to obtain fake news domains. Since one common way previous work has used NewsGuard data is the provided 'trustworthiness' scores for outlets[64–66], we also ran our experiments for trustworthy domains, assessing the effects of a different classification, and showed that our findings are robust to this particular way of filtering for reliable outlets (Table 3).

We relied on NewsGuard annotations because of their particular fit with our research questions, since to the best of our knowledge, the other annotated datasets or manual annotations of news domains do not provide annotations for 'regularly publishing false information'. There is also a broader issue with different domain annotation efforts annotating for different definitions and slightly different phenomena; tackling this challenge is out-of-scope for our work. This challenge along with the issue of list obsolescence complements our broader concerns regarding domain-level categorizations and the need to study articles and individual stories from all domains on their own terms. Our work takes a step in that direction, but as one of the first works to do so, is still somewhat shackled by the precedents set in the field including use of particular resources such as NewsGuard. We support parallel efforts to create publicly available lists for domain and story quality, and our work provides ideas on how to take advantage of existing annotations to move beyond the domain level and identify adjacent news stories from mainstream domains that do not often receive attention in misinformation research. Our method and our codebase can help other researchers take advantage of social media-based co-sharing, while our qualitative case studies can offer journalists and scientists insights into how 'technically factual' information nevertheless plays a role in the broader misinformation landscape.

Obtaining a set of 'mainstream' outlets

We borrowed domain-level ratings from previous work[18]. These ratings are based on Twitter shares of articles (over the 2010 to 2021 period) published by various outlets using the same panel of users as in the overall dataset used in this work (these particular domain-level annotations are made publicly available as part of our codebase). The two criteria we used to demarcate mainstream outlets are detailed below:

(1) Publish predominantly political content. Each domain publishes many articles, but only some of them were deemed as 'political' in domain-level ratings based on the type of user sharing them on Twitter. Using the percentages of all articles published by a domain deemed political content as the political score, 'predominantly political domains' are ones having a score greater than a z-score of 1.0 (greater than the mean plus the standard deviation calculated using all the domain-level political scores).
(2) Popular. Among these domains that predominantly publish political content, we selected the 'popular' ones as those that are (1) in the top 10% for the number of Twitter users sharing a domain's content; and (2) among the top 500 most-visited domains based on the users involved in a YouGov panel survey (2015–2016), with site-visiting statistics derived from passive metering data (based on the analysis and data release in previous work[33]). Using this intersection is a more robust way of deciding the domains that should be deemed 'popular' as opposed to, for example, only considering popularity based on Twitter shares.

Audience-based partisanship scores for news outlets

Our Twitter-based dataset contained a model-based likelihood of identifying as a Democrat at the user level. These scores were used to estimate the average partisan lean of each URL. The scores were normalized to lie within a [−1, 1] (Democrat–Republican) range (in the vein of the traditional 'negative-to-positive' as 'liberal-to-conservative' axis). These Twitter audience-sharing-based scores were aggregated to obtain outlet or domain-level scores following previous work[18]. We used the median of these partisanship scores to classify the outlets as liberal or conservative.

Creating two groups of mainstream news articles

Our process for finding co-shared mainstream articles is visualized in Fig. 1b. We constructed a weighted graph with individual articles or URLs as nodes. Edges between a fake news URL (nodes shown on the left-hand side of Fig. 1b) and a reliable news URL (nodes shown on the right-hand side of Fig. 1b) are weighted by the number of Twitter users in our dataset that shared 'both' the incident node URLs at least once. For example, in Fig. 1b, 8 is the edge weight since eight different users in our panel shared both The Washington Post article and the fake news article displayed. We only considered edges between a pair of nodes where one node is a fake news article and the other node is a reliable article. Our processed dataset contained ~24,000 fake news stories published by 276 domains.

We applied a graph-pruning algorithm[36] to assign each edge a co-sharing score on the basis of the likelihood of both incident node URLs getting shared by the same user, controlling for the individual popularity of each of those URLs in isolation. Specifically, we used the expected likelihood of two articles getting shared by the same group of people, on the basis of how many people shared each of those articles independently, as a null model for determining the actual co-sharing likelihood. For example, an edge assigned a weight of 10, representing 10 people sharing both article 1 and article 2, would receive a higher co-sharing score if article 1 and article 2 each got independently shared by ~20–30 people, compared with them getting shared by 200–300 people each.

Graph-pruning methods have previously been applied to study the media landscape by constructing a domain-level co-exposure network that identifies news outlets with a shared audience[4,67]. However, by focusing on a network of 'individual' news shares, we can investigate

mainstream news stories published by reliable outlets that would otherwise not be considered for their potential role in misinformation circulation on social media. In our implementation of a graph-pruning algorithm on a graph of shared URLs, edges in the top 1% of co-share scores were considered disproportionately co-shared or simply, co-shared (highlighted as the dark green (or darker) edges in Fig. 1b). Following recommendations and practices in previous work[4,36], we set a 'high threshold for deviation from the null model'. To obtain the 0.99 quantile (or top 1%) value to set the threshold, we drew 1,000 random samples of 1,000 × 1,000-sized subgraphs from the entire graph and used the distribution across the samples to estimate the threshold value with 95% confidence. The sampling procedure was used for estimation since using the entire ~24,000 × ~396,000 graph proved to be computationally intractable (where ~24,000 represents number of fake news articles, and ~396,000 represents number of articles from reliable outlets).

We considered articles published by mainstream outlets in this top 1% set of articles as 'the group of mainstream news articles that are co-shared with fake news'. Figure 2 displays examples of top co-shared articles published by The New York Times and The Washington Post, and similarly, Supplementary Fig. 1 displays top co-shared articles published by CNN and Fox News. We note that mainstream outlets form more than 90% of all reliable outlets resulting in ~2,000 domains publishing a total of ~395,000 URLs in our dataset. Stories published by the same set of mainstream outlets that are 'not' in the top 5% of co-sharing likelihood scores form our 'control group'.

Finally, we had ~91,000 news stories in the co-shared group and ~51,000 news stories from the same set of news outlets in our 'control' group, that is, mainstream articles that are not connected with any fake news articles in the top 5% of all likelihood scores. This is a strict control criterion, separating the co-shared and control sets of articles, and explains the lower number of control articles (as opposed to what would be expected when simply considering every other article not in the co-shared set as the comparison group). Apart from these threshold settings of top 1% for co-shared group and never in top 5% for control group, we also ran 18 different experiments with different threshold settings, across top 2%, top 1%, top 0.5% and top 0.1% (for co-shared articles), compared with below 5%, below 4%, below 3%, below 2% and below 1% (for control articles). This is discussed in Supplementary Text 8, with the replicated results across different thresholds presented in Supplementary Tables 27–44. We found that our main results continue to hold for narratives extracted from the content of fake news articles, replicating across the board. There are partial differences for some settings such as using the top 3–1% for the control set, which informs researchers of the sensitivity of our results to some threshold settings: for absolutely no change to our main results, one should use the top 2% or top 1% of co-sharing scores for assigning articles to the co-shared group, and articles never in the top 5% or top 4% of co-sharing scores for assigning articles to the control group.

News article content (headline and story text) of all URLs across the various groups of articles—fake news, co-shared mainstream news and the control group—were automatically extracted. We scraped contents of news articles using the URLs to get the headline, subheading and story texts along with the publication date using the Newspaper3K API, available as a Python library (https://newspaper.readthedocs.io/en/latest/). Our code as well as our entire collection of texts are provided as part of our codebase, although the Twitter data itself will be shared in a restricted manner following Twitter and IRB guidelines.

Narrative extraction

We operationalized narratives using methods and an open-source package ('relatio') developed in previous work[30], available at github.com/relatio-nlp/relatio; we used version 0.2.1 of the relatio package in Python for this study. The conceptual definition used is "sets of relationships between entities that act on each other", which, in language, involves looking at "grammatical statements describing actors, actions and the acted-upon". Figure 3 briefly visualizes the process of extracting the narrative structure through an illustrative example.

Narrative structures extracted consist of the base form:

$$\text{AGENT ENTITY} \rightarrow^{\text{VERB}} \text{PATIENT ENTITY} \in E \times V \times E \quad (1)$$

where E is the set of clustered entities and named entities, and V is the set of all verbs. For example, consider a sentence present in one article in our set of fake news: "Zostavax vaccine caused multiple people to develop shingles." In this case, $\text{vaccine} \rightarrow^{\text{cause}} \text{shingles}$ and $\text{people} \rightarrow^{\text{develop}} \text{shingles}$ could be extracted as these base structures, and we can also have $\text{vaccine} \rightarrow^{\text{cause}} \text{people} \rightarrow^{\text{develop}} \text{shingles}$ (extracted narrative structures need not be limited to triplets). These are then simply represented as strings, or what we call narrative labels or narrative strings ('vaccine cause shingles', 'vaccine cause people develop shingles').

We note that the automated output of this package can be noisy, and not all extracted strings themselves correspond to a meaningful notion of 'who did what to whom'. For this reason, the original paper referred to automatically extracted statements as narrative 'candidates'[30]. We refer to these outputs as narrative strings or narrative labels, and sometimes simply as 'narratives' in this paper, but for fake news articles containing many sentences per article, our work considers 'recurring' narratives as potentially misleading narratives.

As shown in Fig. 3, entities were clustered so that variants of the same entity received the same entity label. Clustering represents the dimensionality reduction step for narrative extraction, since large collections of texts could refer to the same entity in a variety of ways, including grammatical variation ('vaccine' and 'vaccines') as well as various ways to refer to the same named entity ('trump' and 'donald trump'). The pre-specified number of clusters for clustering entities during narrative extraction (Fig. 3) constitutes an important decision. Two settings were used to extract narratives at two levels of granularity from the text data (following ref. 30): 'low-dimensional narratives', allowing fewer, large-sized clusters or latent entities that result in 'broader' narratives, and 'high-dimensional narratives', allowing a higher number of smaller-sized clusters that result in more specific or 'fine-grained' narratives. For example, for the (original, non-processed) sentence "The filing claims that Zostavax caused multiple people to develop shingles.", the method found the low-dimensional narrative 'vaccine cause shingle', and the high-dimensional narrative 'zostavax cause multiple people develop shingle'. Each instance of our hypothesis testing and corresponding results included both these levels of granularity, adding another layer of robustness to our findings. We note that verbs, identified as actions connecting entities in the semantic role labelling step, were not clustered in the narrative extraction method as opposed to entities (subject and patient). This is primarily due to the difficulty of reliable embedding-based clustering for verbs, where verbs implying opposite actions can get clustered together[30]. This had to be avoided since it could lead to opposing narratives (such as 'vaccine harm children' and 'vaccine helps children') being folded under the same narrative. Therefore, all verbs were treated independently and no dimensionality reduction was performed on them[30].

The paper introducing the narrative extraction tool called 'relatio' and open-sourced as a Python package also provides a visualization of the complete algorithm underlying the method in their Fig. 2 (ref. 30). Here we briefly describe the algorithmic process underlying the method that goes from the corpus to extracted narrative statements (these steps are not always sequential and can occur in parallel).

- Step 1: The corpus was split into sentences, and the entire narrative extraction process occurred at the sentence level.
- Step 2: Named entity recognition (NER) was run on the sentences to identify named entities.

- Step 3: Semantic role labelling was used to identify roles for various entities, including agents, patients and attributes. Semantic role labelling also helped identify instances of verb being negated (to retain, for example, 'not winning' as distinct from 'winning', as a verb being negated flips the narrative).
- Step 4: The combination of steps 2 and 3 provided semantic roles with named entities, and roles without named entities.
- Step 5: Sentences (output of step 1) were used to train and fine tune an embedding model (that can work on words and phrases).
- Step 6: The embedding model (step 5) was also trained and fine tuned with the extracted semantic roles without named entities (step 4) to make role embeddings.
- Step 7: The role embeddings (step 6) were used to train *K*-means clustering algorithm and label clusters.
- Step 8: Finally, the extracted roles with named entities (step 4), the verb negations (step 3) and the label clusters (step 7) together resulted in narrative statements or narrative labels or narratives.

Narratives extracted using this tool from speeches on the floor of the US Congress (both the House and the Senate, 1994–2015) have been shown to 'reflect key events in US history', to provide a 'qualitative window' into the priorities and values of US congress members, and to help map ideological disagreements[30]. In addition, the network of entities and the links between them (in the form of verbs or actions) exposed the connections between various topics of debate on the floor of the US Congress. While floor speeches and news articles are different domains of text, the comprehensive range of analyses enabled by this tool in one setting where texts tend to follow some structural similarity (as is the case for floor speeches and for news articles), and crucially, the application of the tool in diverse contexts such as social media posts[30,68], as well as our work's domain of news coverage[69], all together provides assurance for its use as a component in our approach and analysis.

Creating libraries of potentially misleading narratives

In this work, we combined all texts involved in our method: fake news articles, false fact-checked claims, and co-shared and control sets of mainstream news articles, and applied the narrative extraction tool to this combined text data. This ensured that various choices in narrative labels or string representations, such as the chosen entity representation for a cluster, remained consistent across the dataset. We then divided the output according to the different types of data being considered to get the set of narratives for each group of texts.

For potentially misleading narratives, we considered two sources of data. The first consisted of fake news articles (~24,000 URLs), where we considered the headline and story for each article as separate text instances, since many (at least 6 out of 10) people on social media only read the headline of an article[37] (see also americanpressinstitute.org/publications/reports/ survey-research/how-americans-get-news/ and washingtonpost.com/news/the-fix/wp/2014/03/19/ americans-read-headlines-and-not-much-else/). Headlines and story texts were deduplicated and processed to remove information such as domain name mentions and headline text within the story text. This resulted in ~30,000 unique texts derived from fake news articles. The second source consisted of texts of claims that have been fact checked as false (~24,000 unique texts or claims). Potentially misleading narratives from fake news content were defined as recurring narratives: the top 1% in terms of frequency of occurrence (therefore, these narratives occur across several texts in our dataset). We also experimented with all the narratives found in fake news content (Table 2).

For fact-checked false claims, we collected and combined various publicly available datasets. Specifically, sources of false fact-checked claims included (1) claims fact checked by PolitiFact, The Washington Post and FactCheck.org as collected by the Data Commons initiative (datacommons.org/factcheck/download#research-data); (2) fact-checked public health claims[70]; (3) a benchmark dataset of fact-checked claims from PolitiFact[71]; and (4) a large-scale dataset of fact-checked claims coming from 26 different fact-checking websites[72]. All claims are naturally occurring and in the English language. These publicly available claim texts alongside fact-check labels were combined; this combined claims data, along with the process of filtering to retain only false claims, is documented and provided in our codebase.

We used the claims that are labelled as false (or an equivalent label such as 'not true'; our codebase provides all the labels used to create our collection). These claims are provided along with the rest of our data in our publicly available codebase. Since these are statements, usually consisting of a single sentence fact checked as false, we considered 'all' narratives extracted from these data as potentially misleading narratives.

The above process ultimately resulted in three potentially misleading narrative libraries, each containing low- and high-dimensional narratives (Table 2). Table 1 presents examples of potentially misleading narratives.

Hypothesis testing setup

Consider a particular library of potentially misleading narratives and a particular set of high- or low-dimensional narrative labels or strings in that library. Let us denote this particular collection as M, with size $|M|$ and narrative strings $n_1, \ldots, n_i, \ldots, n_{|M|}$. For both groups of mainstream articles (co-shared and control), we obtained the corresponding values of each potentially misleading narrative as its normalized count in the set of all narratives extracted from that particular group of articles. Specifically, let C_s be the collection of all narrative strings present in co-shared mainstream articles (with $|C_s|$ as the total number of unique narrative strings: $s_1, \ldots, s_i, \ldots, s_{|C_s|}$), and C_o be the collection of all narrative strings present in control mainstream articles (with $|C_o|$ as the total number of unique strings: $o_1, \ldots, o_i, \ldots, o_{|C_o|}$). Then, the two lists of values we compared, l_{cs} for co-shared and l_{co} for control, were created by iterating over all narratives (n_i) in M:

$$l_{cs} = \left[\frac{\text{count}(n_i \in C_s)}{\sum_{i=1}^{|C_s|} \text{count}(s_i \in C_s)} \forall n_i \in M \right]$$
$$l_{co} = \left[\frac{\text{count}(n_i \in C_o)}{\sum_{i=1}^{|C_o|} \text{count}(o_i \in C_o)} \forall n_i \in M \right] \quad (2)$$

We then tested whether potentially misleading narratives have a significantly higher presence in co-shared articles ($l_{cs} > l_{co}$; Fig. 1c) using the one-sided Wilcoxon signed-rank test. The same potentially misleading narrative labels (M) were measured for their counts or presence across articles in both groups.

Note on randomization

The only part of our core methodology as developed in this study which involves randomization is the estimation of quantile values (the 0.999, 0.995, 0.99, 0.98, 0.97, 0.96 and 0.95 quantile values) during our graph-pruning-based method of computing co-sharing scores between news from reliable outlets and fake news outlets, to set thresholds for the co-shared and control groups. As specified above, we drew 1,000 random samples of 1,000 × 1,000-sized subgraphs from the entire graph and used the distribution across the samples to estimate the threshold value with 95% confidence. Random sampling in our study was only driven in this particular component by computational feasibility concerns. This random sampling was implemented via the 'random' module or library in Python, which uses the Mersenne Twister as the core generator of pseudo-random numbers[73].

Randomization was not used for any group allocation; hence the concept of blinding was not relevant to this study.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

We used publicly available tweets and the text of published news articles. While our public, no-restriction data sharing ability is limited by the restrictions placed on data sharing by Twitter's Terms of Service, broader privacy concerns regarding the panel of users, and data-use agreements for NewsGuard and TargetSmart data, we made aggregated data files along with all the code required for replicating our findings freely and publicly available on Zenodo at https://doi.org/10.5281/zenodo.15182977 (ref. 74). These data files include all derived data used in our research for testing our hypotheses and everything necessary for replication of the main results (Table 3). The only parts of data not made available in this public, no-restriction setting are: (1) Raw data about the tweets (that is user IDs, tweet IDs, tweet texts and so on), restricted following Twitter policy. While sharing of Twitter data is complicated by Twitter's evolving policy on researcher access, per our interpretation of the latest terms of service from Twitter, tweet IDs can also not be made publicly available. The lack of public availability is also due to user identification concerns for our Twitter panel. (2) NewsGuard-based data that could directly reidentify their annotations (such as groupings of domains, which will directly reveal NewsGuard annotations), in line with legal agreement with NewsGuard (a proprietary data source). (3) US voter file data acquired from TargetSmart in 2017 (with which the Twitter panel was constructed and which ensures the presence of all real users in our data, necessary to investigate our research questions). This proprietary data cannot be directly shared in raw form per legal agreement with TargetSmart. However, Twitter data as well as the data 'derived' from the two proprietary resources used in this research are also made available to researchers who apply for access and sign a data-use agreement, following the precedent set in previous work[4]. This agreement will explicate the requirement that no attempts be made "to identify, reidentify, or otherwise deanonymize the dataset or the restricted resources", and restrict further sharing until explicit approval is obtained from Northeastern University. This is necessary both from a user privacy point of view as per Northeastern's IRB approval for the dataset (17-12-13), as well as legal agreements made with TargetSmart and NewsGuard. Researchers can contact the corresponding author of this study or David Lazer at Northeastern University to obtain access to restricted data. For fact-checked false claims, we collect and combine various publicly available datasets. Specifically, sources of false fact-checked claims include (1) claims fact checked by PolitiFact, The Washington Post, and FactCheck.org as collected by the Data Commons initiative (datacommons.org/factcheck/download#research-data); (2) fact-checked public health claims[70]; (3) a benchmark dataset of fact-checked claims from PolitiFact[71]; and (4) a large-scale dataset of fact-checked claims coming from 26 different fact-checking websites[72]. All claims are naturally occurring and in the English language. These publicly available claim texts alongside fact-check labels are combined; this combined claims data, along with the process of filtering to retain only false claims, is documented and provided in our publicly available data and codebase.

Code availability

All of our code, along with the aggregated dataset files needed to replicate our main findings in Table 3, is available on Zenodo at https://doi.org/10.5281/zenodo.15182977 (ref. 74) under the Creative Commons Attribution 4.0 International licence. Custom code using Python and R was used in all analyses.

References

1. Lazer, D. M. et al. The science of fake news. *Science* **359**, 1094–1096 (2018).
2. Che, X., Metaxa-Kakavouli, D. & Hancock, J. T. Fake news in the news: an analysis of partisan coverage of the fake news phenomenon. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* 289–292 (Association for Computing Machinery, 2018).
3. Weeks, B. E. & Gil de Zúñiga, H. What's next? Six observations for the future of political misinformation research. *Am. Behav. Sci.* **65**, 277–289 (2021).
4. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 US presidential election. *Science* **363**, 374–378 (2019).
5. *Newsguard: Global Leader in Information Reliability* (NewsGuard, 2021); https://www.newsguardtech.com/
6. Allcott, H. & Gentzkow, M. Social media and fake news in the 2016 election. *J. Econ. Perspect.* **31**, 211–36 (2017).
7. Budak, C. What happened? The spread of fake news publisher content during the 2016 US presidential election. In *The World Wide Web Conference* 139–150 (Association for Computing Machinery, 2019).
8. Robertson, R. E. et al. Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature* **618**, 342–348 (2023).
9. Guess, A., Nagler, J. & Tucker, J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* **5**, eaau4586 (2019).
10. Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
11. Aral, S. & Eckles, D. Protecting elections from social media manipulation. *Science* **365**, 858–861 (2019).
12. Lovari, A. Spreading (dis) trust: Covid-19 misinformation and government intervention in Italy. *Media Commun.* **8**, 458–461 (2020).
13. Ognyanova, K., Lazer, D., Robertson, R. E. & Wilson, C. Misinformation in action: fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harv. Kennedy Sch. Misinformation Rev.* https://doi.org/10.37016/mr-2020-024 (2020).
14. Kubin, E. & von Sikorski, C. The role of (social) media in political polarization: a systematic review. *Ann. Int. Commun. Assoc.* **45**, 188–206 (2021).
15. Green, J., Hobbs, W., McCabe, S. & Lazer, D. Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election. *Proc. Natl Acad. Sci. USA* **119**, e2115900119 (2022).
16. Giglietto, F., Marino, G., Mincigrucci, R. & Stanziano, A. A workflow to detect, monitor, and update lists of coordinated social media accounts across time: the case of the 2022 Italian election. *Soc. Media Soc.* **9**, 20563051231196866 (2023).
17. Thorson, K. & Wells, C. Curated flows: a framework for mapping media exposure in the digital age. *Commun. Theor.* **26**, 309–328 (2015).
18. Green, J. et al. Curation bubbles. *Am. Polit. Sci. Rev*. https://doi.org/10.1017/S0003055424000984 (2025).
19. González-Bailón, S. et al. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* **381**, 392–398 (2023).
20. Watts, D. J., Rothschild, D. M. & Mobius, M. Measuring the news and its impact on democracy. *Proc. Natl Acad. Sci. USA* **118**, e1912443118 (2021).
21. Sharma, K., Zhang, Y. & Liu, Y. Covid-19 vaccine misinformation campaigns and social media narratives. *Proc. Int. AAAI Conf. Web Soc. Media* **16**, 920–931 (2022).
22. Allen, J., Watts, D. J. & Rand, D. G. Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science* **384**, eadk3451 (2024).
23. Prochaska, S. et al. Mobilizing manufactured reality: how participatory disinformation shaped deep stories to catalyze action during the 2020 US presidential election. *Proc. ACM Hum. Comput. Interact.* **7**, 1–39 (2023).

24. Starbird, K., DiResta, R. & DeButts, M. Influence and improvisation: participatory disinformation during the 2020 US election. *Soc. Media Soc.* **9**, 20563051231177943 (2023).
25. Wirtschafter, V. et al. Detecting misinformation: identifying false news spread by political leaders in the Global South. *J. Quant. Descr. Digit. Media* https://doi.org/10.51685/jqd.2024.007 (2024).
26. Rosenberg, H., Syed, S. & Rezaie, S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the covid-19 pandemic. *Can. J. Emerg. Med.* **22**, 418–421 (2020).
27. Ebermann, D. in *From Uncertainty to Confidence and Trust* (eds Schmied, J. et al.) 51–64 (Cuvillier Verlag, 2022).
28. Berlinski, N. et al. The effects of unsubstantiated claims of voter fraud on confidence in elections. *J. Exp. Polit. Sci.* **10**, 34–49 (2023).
29. Sloman, S. A. & Lagnado, D. Causality in thought. *Annu. Rev. Psychol.* **66**, 223–247 (2015).
30. Ash, E., Gauthier, G. & Widmer, P. Relatio: text semantics capture political and economic narratives. *Polit. Anal.* **32**, 115–132 (2023).
31. Piper, A., So, R. J. & Bamman, D. Narrative theory for computational narrative understanding. In *Proc. 2021 Conference on Empirical Methods in Natural Language Processing* (eds Moens, M.-F. et al.) 298–311 (Association for Computational Linguistics, 2021).
32. Egelhofer, J. L. & Lecheler, S. Fake news as a two-dimensional phenomenon: a framework and research agenda. *Ann. Int. Commun. Assoc.* **43**, 97–116 (2019).
33. Guess, A. M. (Almost) Everything in moderation: new evidence on Americans' online media diets. *Am. J. Polit. Sci.* **65**, 1007–1022 (2021).
34. Hughes, A. G. et al. Using administrative records and survey data to construct samples of tweeters and tweets. *Public Opin. Q.* **85**, 323–346 (2021).
35. Shugars, S. et al. Pandemics, protests, and publics: demographic activity and engagement on Twitter in 2020. *J. Quant. Descr. Digit. Media* https://doi.org/10.51685/jqd.2021.002 (2021).
36. Dianati, N. Unwinding the hairball graph: pruning algorithms for weighted complex networks. *Phys. Rev. E* **93**, 012304 (2016).
37. Gabielkov, M., Ramachandran, A., Chaintreau, A. & Legout, A. Social clicks: what and who gets read on Twitter? In *Proc. 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* 179–192 (Association for Computing Machinery, 2016).
38. Wilcoxon, F. in *Breakthroughs in Statistics* (eds Kotz, S. & Johnson, N. L.) 196–202 (Springer, 1992).
39. Woolson, R. F. in *Wiley Encyclopedia of Clinical Trials* (Wiley, 2007); https://doi.org/10.1002/9780471462422.eoct979
40. Pratt, J. W. Remarks on zeros and ties in the Wilcoxon signed-rank procedures. *J. Am. Stat. Assoc.* **54**, 655–667 (1959).
41. Dor, D. On newspaper headlines as relevance optimizers. *J. Pragmat.* **35**, 695–721 (2003).
42. Ecker, U. K., Lewandowsky, S., Chang, E. P. & Pillai, R. The effects of subtle misinformation in news headlines. *J. Exp. Psychol. Appl.* **20**, 323–335 (2014).
43. Molek-Kozakowska, K. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse Commun.* **7**, 173–197 (2013).
44. Chen, Y., Conroy, N. J. & Rubin, V. L. Misleading online content: recognizing clickbait as "false news". In *Proc. 2015 ACM on Workshop on Multimodal Deception Detection* 15–19 (Association for Computing Machinery, 2015).
45. Gabriel, S. et al. Misinfo reaction frames: reasoning about readers' reactions to news headlines. In *Proc. 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)* (eds Muresan, S. et al.) 3108–3127 (Association for Computational Linguistics, 2022).
46. Chesney, S., Liakata, M., Poesio, M. & Purver, M. Incongruent headlines: yet another way to mislead your readers. In *Proc. 2017 EMNLP Workshop: Natural Language Processing Meets Journalism* (eds Popescu, O. & Strapparava, C.) 56–61 (Association for Computational Linguistics, 2017).
47. Grice, H. P. in *Speech Acts* (eds Cole, P. & Morgan, J. L.) 41–58 (Brill, 1975).
48. Rogers, K. The latest disinformation threat online? Old news stories. *CBC News* (13 June 2019).
49. Baum, M. A. & Groeling, T. Shot by the messenger: partisan cues and public opinion regarding national security and war. *Polit. Behav.* **31**, 157–186 (2009).
50. Hochschild, A. R. *Strangers in Their Own Land: Anger and Mourning on the American Right* (The New Press, 2016).
51. Polletta, F. & Callahan, J. Deep stories, nostalgia narratives, and fake news: storytelling in the Trump era. *Am. J. Cult. Sociol.* **5**, 392–408 (2017).
52. Williams, D. The marketplace of rationalizations. *Econ. Philos.* **39**, 99–123 (2023).
53. Altay, S., Hacquin, A.-S. & Mercier, H. Why do so few people share fake news? It hurts their reputation. *New Media Soc.* **24**, 1303–1324 (2022).
54. Budak, C. et al. Bursts of contemporaneous publication among high- and low-credibility online information providers. *New Media Soc.* **27**, 726–746 (2023).
55. Vargo, C. J., Guo, L. & Amazeen, M. A. The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016. *New Media Soc.* **20**, 2028–2049 (2018).
56. Tsfati, Y. et al. Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Ann. Int. Commun. Assoc.* **44**, 157–173 (2020).
57. Guo, L. & Vargo, C. "Fake news"' and emerging online media ecosystem: an integrated intermedia agenda-setting analysis of the 2016 US presidential election. *Commun. Res.* **47**, 178–200 (2020).
58. Starbird, K. Facts, frames, and (mis)interpretations: understanding rumors as collective sensemaking. *Center for an Informed Public* https://www.cip.uw.edu/2023/12/06/rumors-collective-sensemaking-kate-starbird/ (Univ. Washington, 2023).
59. Pouliquen, B., Steinberger, R. & Best, C. Automatic detection of quotations in multilingual news. In *Proc. Recent Advances in Natural Language Processing* 487–492 (Incoma, 2007).
60. Harry, J. C. Journalistic quotation: reported speech in newspapers from a semiotic-linguistic perspective. *Journalism* **15**, 1041–1058 (2014).
61. Bergler, S. in *Computing Attitude and Affect in Text: Theory and Applications* (eds Shanahan, J. G. et al.) 11–22 (Springer, 2006).
62. Allcott, H., Gentzkow, M. & Yu, C. Trends in the diffusion of misinformation on social media. *Res. Polit.* https://doi.org/10.1177/2053168019848554 (2019).
63. Lin, H. et al. High level of correspondence across different news domain quality rating sets. *PNAS Nexus* **2**, pgad286 (2023).
64. Guess, A., Aslett, K., Tucker, J., Bonneau, R. & Nagler, J. Cracking open the news feed: exploring what US Facebook users see and share with large-scale platform data. *J. Quant. Descr. Digit. Media* https://doi.org/10.51685/jqd.2021.006 (2021).
65. Edelson, L. et al. Understanding engagement with U.S. (mis) information news sources on Facebook. In *Proc. 21st ACM Internet Measurement Conference* 444–463 (Association for Computing Machinery, 2021).
66. Lühring, J., Metzler, H., Lazzaroni, R. M., Shetty, A. & Lasser, J. Best practices for source-based research on misinformation and news trustworthiness using NewsGuard. *J. Quant. Descr. Digit. Media* https://doi.org/10.51685/jqd.2025.003 (2025).

67. Majó-Vázquez, S., Nielsen, R. K. & González-Bailón, S. The backbone structure of audience networks: a new approach to comparing online news consumption across countries. *Polit. Commun.* **36**, 227–240 (2019).
68. Sipka, A., Hannak, A. & Urman, A. Comparing the language of QAnon-related content on Parler, Gab, and Twitter. In *Proc. 14th ACM Web Science Conference 2022* 411–421 (Association for Computing Machinery, 2022).
69. Ottonello, P. & Song, W. *Financial Intermediaries and the Macroeconomy: Evidence from a High-frequency Identification* Technical Report (National Bureau of Economic Research, 2022).
70. Kotonya, N. & Toni, F. Explainable automated fact-checking for public health claims. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (eds Webber, B. et al.) 7740–7754 (Association for Computational Linguistics, 2020).
71. Wang, W. Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection. In *Proc. 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (eds Barzilay, R. & Kan, M.-Y.) 422–426 (Association for Computational Linguistics, 2017).
72. Augenstein, I. et al. MultiFC: a real-world multi-domain dataset for evidence-based fact checking of claims. In *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (eds Inui, K. et al.) 4685–4697 (Association for Computational Linguistics, 2019).
73. Matsumoto, M. & Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* **8**, 3–30 (1998).
74. Goel, P. Using co-sharing to identify use of mainstream news for promoting potentially misleading narratives (Code + Aggregated Data Files). *Zenodo* https://doi.org/10.5281/zenodo.15182977 (2025).

Acknowledgements

We thank A. Hoyle, P. Rankel, Y. Ting, S. McCabe, B. Swire-Thompson, S. Shugars, K. Joseph, N. Grinberg and A. Mislove for helpful technical support and feedback; A. Smith, A. Uslu, A. Quintana Mathé, A. Wan, B. Ozturan, S. Kumar and Y. Yang for support with manual annotation and constructive discussions. This research was enabled by material resources provided by NewsGuard and TargetSmart. Funding was provided by National Science Foundation award 2008761 (P.G., P.S.R.). National Science Foundation award 2031736 (P.G., P.S.R.), Hewlett Foundation grant number 2019-9224 (D.L.), and Volkswagen Foundation grant 'Reclaiming Individual Autonomy and Democratic Discourse Online' (D.L.) The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

P.G., J.G., D.L. and P.S.R. conceptualized the project. P.G. developed the methodology and software, and conducted formal analysis. P.G. and J.G. conducted investigation. P.G. and P.S.R. performed visualization. J.G., D.L. and P.S.R. supervised the project. P.G. wrote the original paper draft. P.G., J.G., D.L. and P.S.R. reviewed and edited the paper. D.L. and P.S.R. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02223-4.

Correspondence and requests for materials should be addressed to Pranav Goel.

Peer review information *Nature Human Behaviour* thanks Jonathan Bright, Fabio Giglietto and Hernan Makse for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

nature portfolio

Corresponding author(s): Pranav Goel

Last updated by author(s): Apr 11, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The data for this study comes from a pre-existing panel of Twitter users matched to a commercial voter file (originally described in Hughes, et al (2021): Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. Public Opinion Quarterly 85(S1): 323-346.). Details about constructing and validating this dataset for demographic representation of Twitter's U.S. user base are provided in this prior work. The commercial US voter file was obtained from TargetSmart in October 2017. Twitter API v1 was used to collect Twitter data. News article text data was collected using the Newspaper 3k API (version 0.2.8) in Python as a Python library (https://newspaper.readthedocs.io/en/latest/).
Data analysis	Custom code using Python (version >= 3.7.13) and R (version 4.0) was used in all analyses. These programming languages use existing libraries or software packages which enable analyses. All the required packages are listed in our code repository. The most prominent existing library used as a tool in this work is called relatio, publicly available at https://github.com/relatio-nlp/relatio; we used version 0.2.1 of relatio.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Twitter dataset used in this study was previously approved for research (Hughes 2021, Shugars 2021) and collection by the Institutional Review Board (IRB) at Northeastern University (17-12-13).

We use publicly available tweets and the text of published news articles. While our public, no-restriction data sharing ability is limited by the restrictions placed on data sharing by Twitter's Terms of Service, broader privacy concerns regarding the panel of users, and data-use agreements for NewsGuard and TargetSmart data, we make aggregated data files along with all the code required for replicating our findings freely and publicly available at zenodo.org/records/14320676. These data files include all derived data used in our research for testing our hypotheses and everything necessary for replication of main results. The only parts of data not made available in this public, no-restriction setting are:
1. Raw data about the tweets (i.e. user IDs, tweet IDs, tweet texts, etc.), restricted following Twitter policy. While sharing of Twitter data is complicated by Twitter's evolving policy on researcher access, per our interpretation of latest terms of service from Twitter, tweet IDs can also not be made publicly available. The lack of public availability is also due to user identification concerns for our Twitter panel.
2. NewsGuard-based data that could directly reidentify their annotations (such as groupings of domains, which will directly reveal NewsGuard annotations), in line with legal agreement with NewsGuard (a proprietary data source).
3. US voter file data acquired from Targetsmart in 2017 (with which the Twitter panel was constructed and which ensures the presence of all real users in our data, necessary to investigate our research questions). This proprietary data cannot be directly shared in raw form per legal agreement with Targetsmart.

However, Twitter data as well as the data derived from the two proprietary resources used in this research are also made available to researchers who apply for access and sign a data use agreement, following the precedent set in Grinberg et al. (2019). This agreement will explicate the requirement that no attempts be made "to identify, reidentify, or otherwise deanonymize the dataset or the restricted resources", and restrict further sharing until explicit approval is obtained from Northeastern University. This is necessary both from a user privacy point of view as per Northeastern's IRB approval for the dataset (17-12-13), as well as legal agreements made with TargetSmart and NewsGuard. Researchers can contact the corresponding author of this study or David Lazer at Northeastern University to obtain access to restricted data.

For fact-checked false claims, we collect and combine various publicly available datasets. Specifically, sources of false fact-checked claims include {a) claims fact-checked by PolitiFact, Washington Post, and FactCheck.org as collected by the Data Commons initiative (datacommons.org/factcheck/download\#research-data); b) fact-checked public health claims (Kotonya et al. 2020); c) a benchmark dataset of fact-checked claims from PolitiFact (Wang et al. 2017); and d) a large-scale dataset of fact-checked claims coming from 26 different fact-checking websites (Augenstein et al. 2019). All claims are naturally occurring and in the English language. These publicly available claim texts alongside fact-check labels are combined; this combined claims data, along with the process of filtering to retain only false claims, is documented and provided in our publicly available data and codebase.

N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on twitter during the 2016 us presidential election, Science 363, 374–378 (2019).
A. G. Hughes, et al., Using administrative records and survey data to construct samples of tweeters and tweets, Public Opinion Quarterly 85, 323–346 (2021).
S. Shugars, et al., Pandemics, protests, and publics: Demographic activity and engagement on twitter in 2020, Journal of Quantitative Description: Digital Media 1 (2021).
N. Kotonya, F. Toni, Explainable automated fact-checking for public health claims, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) pp. 7740–7754 (2020).
W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) pp. 422–426 (2017).
I. Augenstein, et al., Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) pp. 4685–4697 (2019).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We do not conduct any sex or gender-specific analyses because we do not have theoretical reasons to suspect that our results differ by user sex or gender.
Reporting on race, ethnicity, or other socially relevant groupings	The primary socially relevant grouping we use in our analyses here is partisan identification. For this, we use a model-based estimate of user partisanship, representing the probability of the user identifying as a Democrat, included with our voter file. Where applicable, we control for the average partisanship of URLs as a potential confounder for misinformation narratives.
Population characteristics	See above.
Recruitment	Users were not recruited directly; rather, they were initially identified via the Twitter Decahose and exact-matched to a commercial voter file by first name, last name, and location. Details on the representativeness of this panel relative to other recruitment methods and the demographics of Twitter users as a whole can be found in Hughes, et al (2021): Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. Public Opinion Quarterly 85(S1): 323-346.
Ethics oversight	This research was approved by the Institutional Review Board at Northeastern University (#17-12-13)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a quantitative study, primarily combining natural language processing and network analysis. Linear regression is used in supplemental analyses for robustness checks.
Research sample	The research sample for this study comes from a pre-existing panel of Twitter users matched to a commercial voter file (originally described in Hughes, et al (2021): Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. Public Opinion Quarterly 85(S1): 323-346.). Details about constructing and validating this dataset for demographic representation of Twitter's U.S. user base are provided in this prior work. Specifically, our panel slightly over-represents white and female Twitter users, but is otherwise demographically similar to samples of Twitter users collected via high-quality surveys. This means that our Twitter panel dataset is approximately demographically representative of U.S. based Twitter users. This existing panel dataset is a good fit for our study as our research sample because in this study we are chiefly concerned with (mis)information shared on social media, for which the sharing behavior of users is necessary and the ability to verify that these users are real people is extremely useful. The panel used in our dataset was constructed by identifying users active in the Twitter Decahose data (a random 10% sample of all of Twitter that was available to academics via the Twitter API before 2022) with a unique first name/last name/location that also appeared in a commercial voter file. In other words, Twitter users were exact-matched to a commercial voter file by first name, last name, and location. All matches were retained in the panel (i.e. we did not randomly sample). After the user panel was constructed, users' tweets were collected at regular intervals using the Twitter API. Further details about construction and validation of this dataset are provided in Hughes, et al (2021): Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. Public Opinion Quarterly 85(S1): 323-346. We retain English tweets that share news articles (published by English-language information sources or outlets per NewsGuard) from May 1 2018 to November 14 2021. This results in an initial dataset containing ~14.6 million unique URLs published by ~4000 domains shared by ~600,000 different users. For each tweet, we have the user ID of the author, and the news URL being shared in the tweet. We retain URLs that were shared by at least 20 different users (to control for sparsity effects; we did not conduct further sampling as our study is large-scale observational), resulting in ~420,000 different URLs published by ~2400 domains shared by ~450,000 different users. Compared to the full panel of about 1.6 million U.S. Twitter users, our subset of ~450,000 users slightly overrepresents white and female Twitter users, and overrepresents users registered as Democrats. Our subsample also slightly underrepresents Twitter users in the 18-29 age bucket and, in almost equal proportion, slightly overrepresents users in the 50-64 age bucket.
Sampling strategy	The sample was constructed by identifying users active in the Twitter Decahose with a unique first name/last name/location that also appeared in a commercial voter file, described in greater detail in Hughes, et al (2021): Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. Public Opinion Quarterly 85(S1): 323-346. All matches were retained in the panel (i.e. we did not randomly sample). We considered all URL sharing within the examined time period (May 2018 to November 2021), where an URL has to be shared by at least twenty different users. Therefore, the construction of our underlying dataset and our core study does not involve sampling as our study is large-scale observational.
Data collection	After the user panel was constructed, users' tweets were collected at regular intervals using the Twitter API. No one was present besides the participants in the data collection, and the researcher collecting this original dataset was blinded to the study's experimental condition as well as hypothesis.
Timing	We consider URLs shared between May 2018 and November 2021.
Data exclusions	To the best of our understanding, no data was excluded from the analyses.
Non-participation	As subjects were not recruited (their data was passively collected as our study is large-scale observational), no subjects declined to participate or opted out of the research.
Randomization	Our main quantitative experiment involves comparing two sets of news articles for the presence of certain narrative structures, and does not involve covariates. The only part of our core methodology as developed in this study which involves randomization is the estimation of quantile values (the 0.99, 0.98, 0.97, 0.96, 0.95 quantile values) during our graph-pruning based method of computing co-sharing scores between news from reliable outlets and fake news outlets, in order to set thresholds for the co-shared and control groups. We draw 1000 random samples of 1000 × 1000 sized subgraphs from the entire graph and use the distribution across the samples to estimate the threshold value with 95% confidence. The sampling procedure is used for estimation since using the entire ~ 24, 000× ~ 396, 000 graph was computationally intractable (where ~ 24, 000 represents number of fake news articles, and ~ 396, 000 represents number of articles from reliable outlets). Random sampling in our study is only driven in this particular component by computational feasibility concerns. This random sampling is implemented via random module or library in Python, which uses the Mersenne Twister as the core generator of pseudo-random numbers (Matsumoto et al. 1998).Randomization is not being used for any group allocation, and therefore, the concept of blinding is also not relevant to this study.

M. Matsumoto, T. Nishimura, Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation (TOMACS) 8, 3–30 (1998).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A