Contextualizing Misinformation: A User-Centric Approach to Linguistic and Topical Patterns in News Consumption

ROSS DAHLKE*, School of Journalism and Mass Communication, University of Wisconsin-Madison, United States
FANGJING TU*, Social Media Lab, Stanford University, United States
YI-CHIA WANG*, Social Media Lab, Stanford University, United States
YINGDAN LU, Department of Communication Studies, Northwestern University, United States
BLAIN W ENGEDA, Social Media Lab, Stanford University, United States
JEFFREY T. HANCOCK, Social Media Lab, Stanford University, United States

Exposure to misinformation poses significant challenges to democratic processes and public health, particularly during critical events like elections. This study adopts a user-centric approach to analyze the linguistic features of misinformation actually consumed by individuals during web browsing. Using data from a nationally representative panel of 1,240 American adults and their web-browsing data (21M URL visits) during the 2020 U.S. Presidential Election, we examine linguistic and topical differences in the content of 91K unique misinformation and hard news webpages by utilizing natural language processing techniques and Large Language Models. We find that misinformation consumed by users is generally easier to read, exhibits higher negative sentiment, and employs more moral language than hard news. We also find significant linguistic variations across topics—misinformation can be diverse and vary in linguistic features depending on the subject matter. We also identify heterogeneity across key user characteristics: older adults consume more misinformation about COVID-19 and health, with content showing more negative sentiment and fewer moral terms than expected. Republicans engage with misinformation characterized by more negative sentiment and higher moral language, focusing less on health topics and more on social and political issues. These results highlight the importance of a user-centric approach and suggest that interventions to combat misinformation should be tailored to specific topics and user characteristics for greater effectiveness.

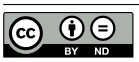
CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Social and professional topics** → **User characteristics**.

Additional Key Words and Phrases: digital trace data, LLMs, demographic targeting, health misinformation, content analysis, U.S. Presidential Election, intervention design, older adults

ACM Reference Format:
Ross Dahlke*, Fangjing Tu*, Yi-Chia Wang*, Yingdan Lu, Blain W Engeda, and Jeffrey T. Hancock. 2025. Contextualizing Misinformation: A User-Centric Approach to Linguistic and Topical Patterns in News

* The authors contributed equally to this work.

Authors' Contact Information: Ross Dahlke*, School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, United States; Fangjing Tu*, Social Media Lab, Stanford University, Stanford, California, United States; Yi-Chia Wang*, Social Media Lab, Stanford University, Stanford, California, United States; Yingdan Lu, Department of Communication Studies, Northwestern University, Evanston, Illinois, United States; Blain W Engeda, Social Media Lab, Stanford University, Stanford, California, United States; Jeffrey T. Hancock, Social Media Lab, Stanford University, Stanford, California, United States.



ACM 2573-0142/2025/11-ARTCSCW390
https://doi.org/10.1145/3757571

Consumption. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW390 (November 2025), 40 pages. https://doi.org/10.1145/3757571

1 Introduction

Tens of millions of adults in the United States alone are exposed to misinformation—information that is false, misleading, or unsubstantiated [124]—during election seasons [64, 109]. Misinformation exposure can affect people's beliefs and attitudes, including on the legitimacy of democratic elections [18, 37], vaccination intentions [6], and climate change [195]. The World Economic Forum even recently listed misinformation as society's most concerning problem [181]. As a result, a burgeoning area of human-computer interaction research is focused on combating misinformation, specifically through the design of interventions to help people identify misinformation, e.g., [2, 9, 65, 72, 73, 127, 190].

A growing body of intervention research encourages people to pay attention to manipulative language to help discern between misinformation and credible information. For example, digital literacy and prebunking interventions often forewarn individuals about how misinformation may attempt to manipulate people's emotions and use of negative emotional language, such as anger [87, 150, 151]. While these interventions have been found effective in helping people identify misinformation in controlled experimental settings, it remains unclear whether the misinformation people encounter in their real-life news consumption contains these linguistic patterns.

Past research into linguistic differences between misinformation and hard news has compared content created by producers of misinformation, such as abcnews.com.co (a misinformation website that attempts to mimic the legitimate abcnews.com), with content produced by hard news organizations, such as *The New York Times*. One important study [28], for example, found that misinformation has higher readability, lower sentiment, higher use of moral language, and lower perplexity. While this approach provides a comprehensive assessment of how misinformation purveyors produce content with different linguistic characteristics than hard news, it does not consider whether people actually consume this information. Indeed, past work has found that the types of articles that misinformation producers primarily create are very different from the types of articles that individuals actually visit [107].

In this paper, we adopt a user-centric approach that examines the linguistic differences in content consumption based on visits to misinformation and hard news websites, drawing on data from a nationally representative panel of 1,240 American adults collected in the months around the 2020 U.S. Presidential Election. When we adopt a user-centric approach, we can explore several new questions to advance our understanding of the content of misinformation within its naturalistic context, rather than relying on researcher-selected examples [42]. First, we can analyze which types of misinformation topics are more likely to be consumed in relation to hard news, such as health, politics, or social issues. As Large Language Models (LLMs) have demonstrated significant potential in enhancing large-scale content and topic analysis through their generative and reasoning capabilities [32, 57, 113, 145, 160] and have been applied to misinformation detection and mitigation [46, 89, 177], we propose a novel approach that integrates unsupervised machine learning, human-in-the-loop methods, and LLMs to categorize content by topic. Recent research from the deception detection literature suggests that the language of deception can vary substantially across genres and motivations [98]. Indeed, we find significant heterogeneity in the overall linguistic differences between misinformation and hard news across five topics: social issues, U.S. electoral politics, COVID-19, health, and general news. For instance, health misinformation had the largest deviations from high-quality health news across the topics: health misinformation is substantially more readable, more negative, and contains less moral language than hard news.

In addition, the user-centric approach allows us to examine how linguistic patterns in misinformation vary across different kinds of individuals, such as younger vs. older adults or conservatives vs. liberals. Prior research suggests that these groups, for example, consume and share misinformation in very different ways, e.g., [64, 109]. We find that the misinformation older adults consume has more negative sentiment but less moral language than that of younger adults. At the same time, Republicans consume misinformation that has more negative sentiment and more moral language than non-Republicans.

By adopting this user-centric approach to misinformation consumption, this paper makes several key contributions. First, we replicate prior research by confirming that misinformation shows distinct linguistic characteristics compared to hard news. Second, we show that misinformation isn't limited to politics; it also spans topics like health, social issues, and general news, each with its own linguistic patterns. This highlights that misinformation is a broad phenomenon, not just a political one. Third, we find that the linguistic features of misinformation vary not only by topic but also by audience. Individual differences—such as age and political orientation—shape the types of linguistic cues people encounter in misinformation. Together, these findings suggest the importance of designing tailored interventions that support diverse groups by aligning with the specific linguistic patterns they are most likely to encounter in misinformation content.

2 Related Work

2.1 Linguistic Differences Between Misinformation and Hard News

Recent misinformation detection research has made progress in identifying patterns in misinformation. Automated detection methods, including machine learning approaches and network analysis, have reported high accuracy in misinformation detection, with some models exceeding 80% performance on benchmark datasets [3, 66, 135, 155, 157, 194]. However, major challenges remain, as many misinformation detection models are trained on pre-processed datasets that do not reflect the content people actually engage with online. Additionally, evaluation methods often fail to properly separate training and testing data, leading to overestimated accuracy and limited generalizability when applied to real-world misinformation [186]. Furthermore, much of the existing research relies on algorithms that lack interpretability and transparency [81, 167]. Despite these limitations, linguistic cues are crucial indicators for detecting misinformation. The primary assumption behind analyzing linguistic cues in misinformation is that misinformation purveyors may reveal their deception through specific language patterns, either unintentionally or as a result of deliberate attempts to exploit individuals' cognitive and emotional vulnerabilities [102, 167, 175].

Prior misinformation research highlights several cognitive vulnerabilities that contribute to individuals' susceptibility to false information. First, dual-process theories [47, 132, 170] suggest that engaging in intuitive rather than analytical thinking increases the likelihood of believing and sharing misinformation [131, 133, 134]. The overwhelming volume of information online further encourages reliance on cognitive shortcuts to evaluate information, such as emotional reactions (e.g., anger and fear) and social endorsement cues (e.g., number of likes and shares) [100, 169]. Second, truth-default theory suggests that individuals may have a systematic tendency to accept information they encounter as true, and are therefore less likely to cast doubt even when encountering false information [19, 85]. Third, according to identity-based motivated reasoning, when individuals are exposed to information that aligns with their prior beliefs or identity, they may be motivated to reason in a way that selectively accepts misinformation that reinforces their political or social identity [174]. These cognitive vulnerabilities may be further amplified when individuals lack digital media literacy skills, a limitation particularly common among older adults, which can reduce their ability to discern misinformation online [110]. Misinformation creators may exploit

these vulnerabilities by utilizing linguistic styles that appeal to cognitive shortcuts, encouraging intuitive processing through emotional and identity-based cues. However, their effectiveness is often examined in datasets that may not capture the full complexity of how misinformation spreads in real-world scenarios.

There are several important reasons to understand how misinformation differs linguistically from hard news. One is the goal of misinformation detection, and the majority of the analysis of the linguistic features of misinformation attempt to address this goal, e.g., [81, 84, 91, 178, 192]. Other work seeks to understand the linguistic patterns in misinformation to improve our understanding of how people process information and potentially come to believe misinformation [91].

One important study in this vein of research is the linguistic examination of the Fake News Corpus by Carrasco-Farré [28], which identified four key dimensions in which misinformation producers create content that is linguistically distinct from reliable hard news. Misinformation is generally more readable than hard news. Misinformation also tends to have lower perplexity, which is a measurement of lexical diversity that reflects the variety of different words in a text. Lower perplexity means that misinformation often includes a more limited vocabulary, making it less complex and surprising. Readability and perplexity are both measures of cognitive effort, suggesting that, on average, the language used in misinformation makes it easier for readers to process and understand. Misinformation also tends to contain more negative sentiment than hard news, and it contains more moral language. The use of a strongly negative emotional tone and frequent moral terms are key indicators of emotional evocation and the leveraging of group identity and self-categorization mechanisms, which contribute to the rapid and frequent spread of misinformation [23, 165].

The low cognitive processing and increased emotional and moral language content may be one reason that misinformation spreads more widely and is easier for people to believe [91, 131]. In contrast, hard news typically follows journalistic standards of objectivity and complexity, requiring more effortful cognitive processing and the avoidance of emotional language to maintain balanced reporting [118]. In the present study, we first sought to replicate and extend these past results, and we hypothesize that:

H1: Misinformation that was consumed by people during their web-browsing activity during the 2020 election, when compared to visited hard news content, (H1a) is more readable, (H1b) has lower perplexity, (H1c) is higher in negative sentiment, and (H1d) uses more moral language.

2.2 Linguistic Styles Differ by Topic

A general approach to analyzing linguistic features in misinformation versus hard news may obscure important differences in lexical patterns of misinformation across different thematic areas. There are substantial topical variations in online news and misinformation. For example, misinformation and credible news often emphasize different topic areas within COVID-19 coverage [193]. Certain subtopics, such as "virus" and "conspiracy," tend to display a more negative tone compared to others like "vaccine" and "cures" [31].

This observed topical heterogeneity is important given recent research in language and deception detection, such as the Contextual Organization of Language and Deception (CoLD) framework [98]. Developed within an interpersonal context, this framework suggests that contextual factors, such as the genre or topic, influence how language is used in deceptive communication. We extend this understanding to misinformation, arguing that the topic shapes linguistic patterns by constraining them to specific linguistic norms. For instance, health-related misinformation might show more sophisticated language than misinformation on social issues [56], as this topic often adheres to more formal writing conventions of science communication.

Therefore, a one-size-fits-all approach may fail to capture the subtle ways in which language is modified to suit different audiences. Understanding the characteristics of misinformation requires a contextual approach that involves a more nuanced examination of how linguistic features differ by topics rather than relying solely on a universal analysis of linguistic features. In our user-centered approach, we examine the kinds of misinformation topics that our users consumed during their web browsing:

RQ1: How do the linguistic features vary between misinformation and hard news across different news topics?

2.3 Individual Differences in Misinformation Linguistic Styles and Topics

A user-centric approach to misinformation acknowledges that different individuals consume various types of content. For example, older adults exhibit distinct preferences for specific news topics. Older adults prioritize health-related news due to its direct impact on their well-being. During the COVID-19 pandemic, older adults followed COVID-19 news more closely than younger adults [76]. In addition, they are more active in political participation [139] and more likely to engage with news more frequently [21]. As such, we predict that:

H2: Older adults are more likely to consume misinformation in health, COVID-19, and electoral politics than other topics.

Older adults also have distinct cognitive processing characteristics, tending to favor and attend to more positive rather than negative information [147] and finding positive media content to be more meaningful [97]. This difference is known as the age-related positivity effect and is predicted by the socioemotional selectivity theory (SST), which posits that as individuals perceive their time horizon shortening with age, they shift their motivation toward prioritizing physical and emotional well-being [147]. Empirical studies support the prediction that older adults are more likely to experience positivity bias, e.g., [120]. Finally, older adults may prefer content that requires lower cognitive effort, due to age-related cognitive changes [26] and greater perceived cost of cognitive effort [182]. Given these preferences for older adults, we predicted that:

H3: Compared to non-older adults, older adults will consume misinformation involving (a) higher positive sentiment, (b) lower readability, and (c) lower perplexity.

Political conservatives represent another group with distinct media consumption patterns and have been found to encounter more misinformation than liberals [64, 109]. However, the underlying factors driving this disparity remain unclear. One possibility is that partisans are drawn to misinformation about different topics. There could be partisan heterogeneity in the topics of misinformation consumption. This possibility is bolstered by the fact that different partisan supporters believe different conspiracy theories [44, 171]. If there are partisan differences in the topics of consumed misinformation, it could point to answering why this discrepancy exists in topline exposure levels. Thus, we ask:

RQ2a: What misinformation topics are Republicans more likely to consume compared to non-Republicans?

Another possibile explanation for differences in misinformation consumption is that partisans are fundamentally drawn to different types of content due to variations in cognitive style. While there is ongoing debate regarding whether there are differences in cognitive styles between Democrats and Republicans (e.g., [14, 40]), proponents of partisan asymmetry (e.g., [75]) argue that Republicans tend to score higher than Democrats on traits such as dogmatism, need for closure, and preference for structure and order, while scoring lower on uncertainty tolerance, cognitive reflection, and

need for cognition. These characteristics may lead Republicans to rely more on intuition and prefer information that requires less cognitive load and more emotional content compared to Democrats. Given these potential differences, we ask the following research question:

RQ2b: Does the misinformation consumed by Republicans exhibit different linguistic patterns compared to that consumed by non-Republicans?

2.4 Large Language Models (LLMs) and Topic Understanding in Misinformation Context

Previous research has presented various methods to define topics within large text corpora, including dictionary-based approaches, supervised learning, and unsupervised learning techniques [61]. For example, unsupervised probabilistic topic modeling has been widely applied to analyze latent topics in texts across different languages in political communication [88, 148]. However, this approach often requires substantial human validation of both keywords and representative text outputs, and the results, presented as word clusters, are often difficult to interpret [113]. At the same time, traditional topic modeling techniques, such as Latent Dirichlet Allocation (LDA), often depend on word co-occurrence patterns, which may overlook nuanced semantic meanings and contextual language use.

Advances in Large Language Models (LLMs) offer a promising approach to understanding topics in large-scale text data more efficiently. Research has shown that LLMs can perform at a level comparable to human annotators in identifying social, psychological, and political concepts using zero-shot methods [32, 57, 129, 145]. Notably, LLMs can capture contextual meanings, emergent language patterns, and subtle thematic signals, enabling more accurate and detailed topic categorization [113]. With their reasoning and generative capabilities, LLMs enable researchers to generate human-readable topic labels and descriptions, thereby enhancing model interpretability. Furthermore, researchers have found that LLMs can complement traditional topic modeling by serving as an assessment tool for evaluating outputs from conventional methods [160]. Yet, the replicability of topic results by LLMs can be challenging due to their black-box nature, and algorithmic biases may potentially affect topic modeling results.

While LLM-based approaches have been applied across various domains, such as understanding scientific and political information [11, 126], they have also been used to combat and mitigate misinformation. For example, Lucas et al. [89] leverage LLMs' generative and reasoning capabilities to develop a "Fighting Fire with Fire" strategy for identifying and countering both human- and LLM-generated disinformation. Other studies further utilize LLMs to design systems that generate explanations or prioritize accurate information over misinformation. However, very few studies clearly identify topic patterns in user-centric misinformation data compared to other types of information, which is the focus of this research.

3 Method

We administered a two-wave survey and collected web-browsing data (i.e., every URL participants visited) via the survey company YouGov. YouGov collects this web-browsing data via their YouGov Pulse software. Participants (N = 1,240) installed the YouGov Pulse software onto their desktop/laptop computers, smartphones, and tablets. The software then passively collects all URL visits made by the participants in exchange for financial compensation. All participants provided informed consent for our specific research project (this study was approved by the [anonymized] Institutional Review Board) and were compensated.

3.1 Participants

For these analyses, we focus on visits to misinformation or hard news domains (see Section 3.2). After filtering down to participants who visited at least one of these web pages of interest, our sample includes 719 participants. Among these participants, 210 (29.2%) were 65 years or older, 509 (70.8%) were between the ages of 18 and 64. A total of 350 (48.7%) participants identified as male, and 369 (51.3%) identified as female. In terms of race, 589 (81.9%) identified as white, 53 (7.4%) as Black, 32 (4.5%) as Hispanic, and 45 (6.3%) as another race or ethnicity, 226 (31.4%) self-identify as Republicans, while 493 (68.6%) are not Republicans. YouGov provides weights for participants to match a nationally representative sample of American adults, which we use in our analyses.

3.2 Procedures

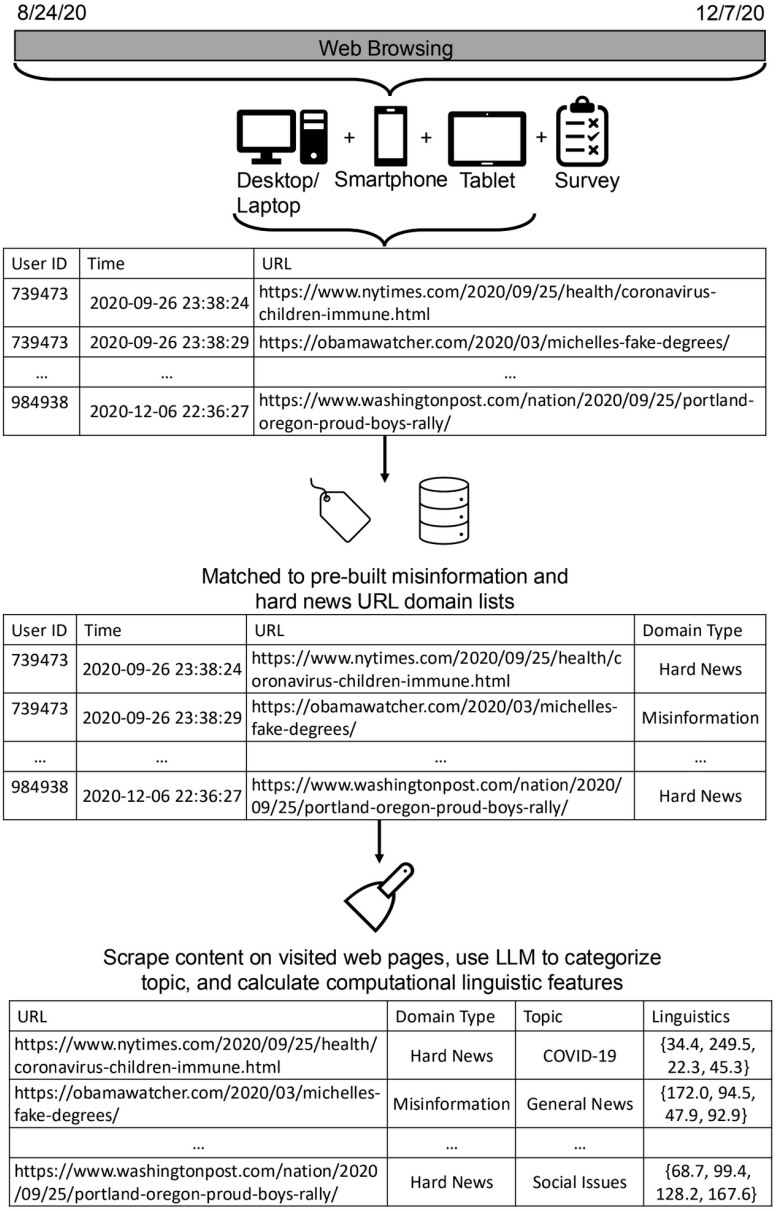
Our study procedures are detailed in Figure 1. After collecting the web-browsing data ($N_{visits} = 21M$), we narrowed our URLs to only hard news and misinformation. Drawing from Garimella et al. [53], who similarly analyzed online news at the domain level, we defined "hard news" as content focused on factual reporting of current events in politics, economics, or current events. For misinformation, we follow Nyhan and Reifler [124], who define it as "information that is false, misleading, or unsubstantiated."

We relied on domain-level classifications from Dahlke et al. [38] to operationalize these definitions. Hard news domains were identified by combining the list from Bakshy et al. [13] with NewsGuard's[1] ratings of 5,471 websites that do not repeatedly publish false content. Misinformation domains were classified using NewsGuard's ratings of sites that "repeatedly publish false content" combined with established academic lists from Guess et al. [64] and Allcott et al. [4], resulting in 1,796 unique domains. Following Dahlke et al. [38], we removed visits to dynamic pages, e.g., homepages, and stripped query parameters from URLs to ensure we captured stable content. We then scraped the remaining URLs using a headless Chrome web browser to extract the article text. This resulted in a dataset of 91,424 unique misinformation and hard news URLs.

3.2.1 LLM-enhanced Topic Labeling. To categorize the content of the scraped web pages, we applied a multi-stage, LLM-enhanced approach to categorizing the topics of misinformation and hard news articles. To leverage the strengths and mitigate the limitations of different topic modeling methods [113, 160], this research combined unsupervised topic modeling with zero-shot LLM approaches. We began by applying unsupervised topic modeling to misinformation articles using the 'stm' R package and identified 20 as the optimal number of topics using the 'searchK' function in 'stm' [148]. Then, two members of the research team independently labeled the topics by examining the top 10 words with the highest topic probability, FREX, and lift. We also analyzed the top five documents with the highest assigned probability to each topic. After reviewing and consolidating similar topics, we established the final four topic categories: *Health, COVID-19, Social Issues, and U.S. Electoral Politics*. This inductive method allowed us to identify the most common content categories of misinformation articles. Additionally, we created a "General News" category for articles that did not fit into any of the predefined topics. A "Missing" category was also added to exclude texts indicating that the page has access denied.

We created a prompt (see Appendix A) based on the top words and documents from these categories to categorize all hard news and misinformation articles. We applied the prompt with GPT-4o mini for categorization. We validated GPT-4o mini's performance on a random subset of 100 articles by comparing GPT outputs and human-labeled categories. Two members of the

[1]NewsGuard is an organization of former professional journalists and news editors who manually rate news websites based on journalistic criteria, including whether they repeatedly publish false content. NewsGuard claims to cover "95%+ of online engagement with news." https://www.newsguardtech.com/



User ID	Time	URL
739473	2020-09-26 23:38:24	https://www.nytimes.com/2020/09/25/health/coronavirus-children-immune.html
739473	2020-09-26 23:38:29	https://obamawatcher.com/2020/03/michelles-fake-degrees/
...	...	...
984938	2020-12-06 22:36:27	https://www.washingtonpost.com/nation/2020/09/25/portland-oregon-proud-boys-rally/

User ID	Time	URL	Domain Type
739473	2020-09-26 23:38:24	https://www.nytimes.com/2020/09/25/health/coronavirus-children-immune.html	Hard News
739473	2020-09-26 23:38:29	https://obamawatcher.com/2020/03/michelles-fake-degrees/	Misinformation
...	...	...	...
984938	2020-12-06 22:36:27	https://www.washingtonpost.com/nation/2020/09/25/portland-oregon-proud-boys-rally/	Hard News

URL	Domain Type	Topic	Linguistics
https://www.nytimes.com/2020/09/25/health/coronavirus-children-immune.html	Hard News	COVID-19	{34.4, 249.5, 22.3, 45.3}
https://obamawatcher.com/2020/03/michelles-fake-degrees/	Misinformation	General News	{172.0, 94.5, 47.9, 92.9}
...	...	...	
https://www.washingtonpost.com/nation/2020/09/25/portland-oregon-proud-boys-rally/	Hard News	Social Issues	{68.7, 99.4, 128.2, 167.6}

Fig. 1. Overview of methodological approach. First, we collected web-browsing data from August 24, 2020, to December 7, 2020, from participants' desktops/laptops, smartphones, and tablets. We also administered a survey to participants to capture demographic and individual characteristics. Next, we categorized websites based on existing lists of misinformation and hard news domains (see Dahlke et al., 2023). Then, we scraped the content of each of these webpages. Finally, we categorized the pages into topics (see Section 3.2.1) and calculated the linguistic features of the content (see Section 3.4).

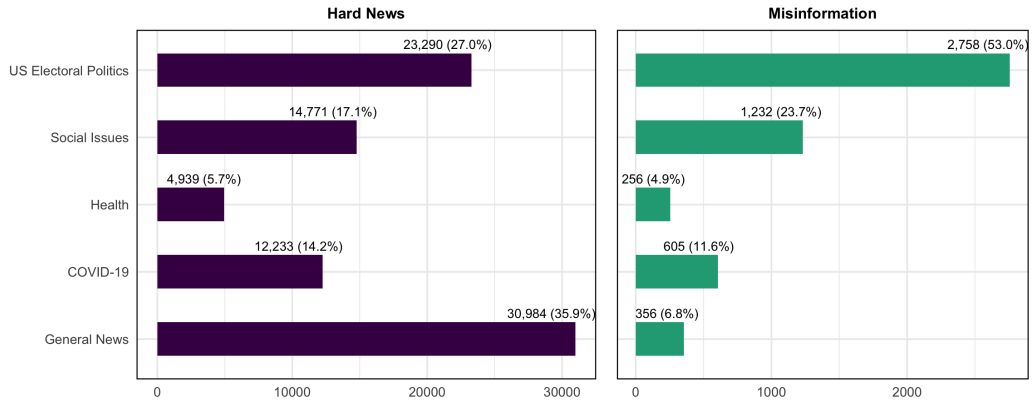


Fig. 2. Topical distribution (%) for hard news (left) and misinformation (right).

research team independently coded the 100 articles across the six categories. The results showed a high level of agreement between human coding and GPT-4o mini coding (Cohen's kappa = 0.76). The topics categorized by GPT-4o mini served as the topical categories used in our analysis, with articles labeled as "Missing" excluded. The frequency distribution of each topic is plotted in Figure 2, and selected case studies for each topic category are shown in Appendix B.

In descriptively and visually examining the temporal distribution of the topics (Figure 3), we observed several notable patterns in both misinformation and hard news consumption across topics from August to December 2020. First, U.S. Electoral Politics showed the most dramatic temporal variation, particularly in misinformation content (Panel A). There was a notable build-up in electoral politics misinformation preceding Election Day 2020, with the percentage rising from around 2-3% pre-election to peaks of nearly 6% of total articles (misinformation plus hard news) just before the election and in the weeks after. Looking at hard news (Panel B), we saw that Electoral politics hard news showed an expected spiked on Election Day, rising to over 60% of content, before declining in the weeks after. General news consistently comprised the highest percentage (30-45%) of content throughout the period. The other topics—health, COVID-19, and social issues—maintained relatively stable levels in hard news coverage, typically ranging between 5-20% of content.

3.3 Measures

Building on prior work on the lexical characteristics of misinformation [28, 81], we measure four linguistic characteristics of misinformation: two reflect the cognitive efforts to process the content (readability and perplexity), and two focus on the emotional aspects of the content (sentiment and morality).

Readability measures the complexity of a text based on sentence length and word syllables. It is a widely used measure in linguistic analysis to detect misinformation [1, 68, 70, 105, 194]. We used the Flesch–Kincaid Grade Level to calculate readability, a metric designed to reflect the level of education required to understand a given text, where higher scores indicate more complex and harder-to-read text [159].

Perplexity is a commonly used metric to measure lexical diversity [58, 60]. It captures the extent to which the words within a text are unpredictable and surprising. Text with higher perplexity indicates more varied word use and greater uncertainty, and lower perplexity means simpler language with more predictable word patterns [16]. Longer sentences, more syllables per word, and greater uncertainty require greater cognitive effort to understand the text. Readability and

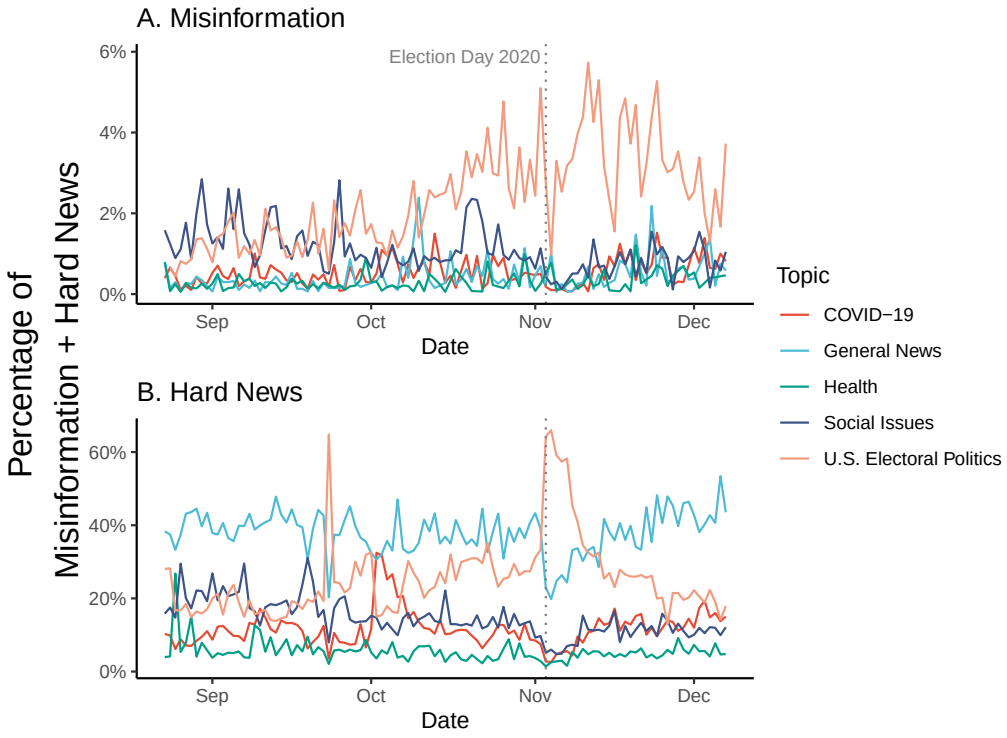


Fig. 3. Temporal Distribution (% of combined) for Misinformation and Hard News Topics. The x-axis represents dates. The y-axis represents the percentage of misinformation or hard news under each topical category. Lines are colored by topic. The vertical dotted line represents U.S. Election Day 2020.

perplexity require a higher cognitive load for information processing. We calculated the perplexity of a text using the GPT-2 model, as it was primarily trained on web-crawled data [141]. This makes it a suitable base model for comparison with the web-browsing data in our analysis [172].

Sentiment analysis captures the emotional polarity of a text, categorizing a text as positive, negative, or neutral. Sentiment plays a significant role in the sharing and acceptance of misinformation, making individuals more vulnerable to false information and resistant to correction attempts [68, 100]. Using the AFINN lexicon dictionary [122], text with more positive words receives higher positive scores on the sentiment scale. In contrast, texts containing more negative words are assigned lower and negative scores. Texts with a more balanced emotional tone will be scored as neutral around 0.

Morality was measured by the frequency of moral terms present in the text, based on the Moral Foundations Dictionary [52, 130]. This validated dictionary includes a list of terms that reflect either support for or violation of core moral foundations. For example, it contains authority-related words such as "obey," "honor," and "disrespect," as well as fairness-related terms like "justice," "equality," and "prejudice." [59]. The measure is sensitive to text length; therefore, we calculated morality as the number of moral terms per 100 words. This normalization enables meaningful comparisons across texts of varying lengths. A higher score indicates that the text includes more morality terms and appeals more to intuitions and values.

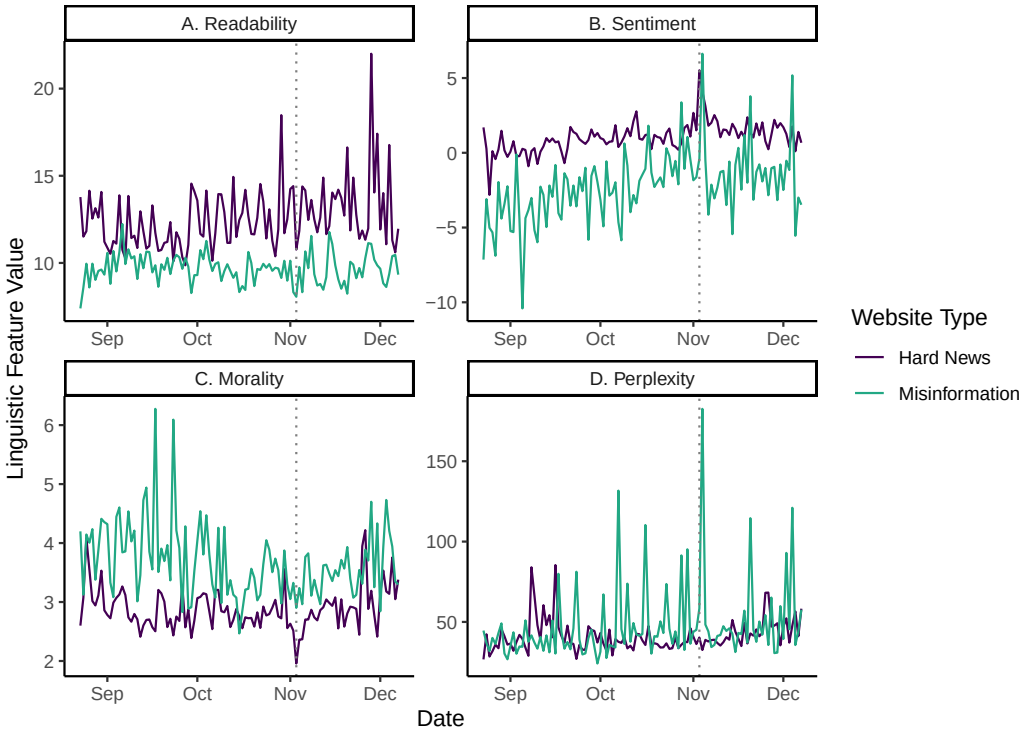


Fig. 4. Temporal Distribution (average of measures) for Misinformation and Hard News linguistic features. The x-axis represents dates. The y-axis represents the average linguistic measure. Lines are colored by hard news or misinformation. The vertical dotted line represents U.S. Election Day 2020.

In examining the temporal patterns of linguistic features (Figure 4) [153], we observe several notable trends across the study period from August to December 2020. Panel A shows that misinformation consistently maintains lower readability scores (indicating easier-to-read content) than hard news throughout the period, with this gap widening slightly around Election Day. Panel B reveals that misinformation exhibits consistently more negative sentiment than hard news, with both types of content showing increased volatility in the weeks surrounding the election. Panel C demonstrates that misinformation generally contains more moral language than hard news, though this difference diminishes somewhat in the post-election period. Finally, Panel D shows considerable fluctuation in perplexity scores for both content types, with misinformation displaying greater variance and several notable spikes, particularly around early November. These temporal patterns suggest that while the linguistic differences between misinformation and hard news remain relatively stable, there may be variations around key events like elections.

Age was measured through a recurring survey administered to participants in the YouGov panel. Respondents were asked, "In what year were you born?" and prompted to enter their birth year. This approach is generally more robust than asking for age directly, as it reduces the likelihood of vague or inaccurate responses. Additionally, YouGov updates this information every 27 months. For participants who remain in the panel for extended periods, multiple records of birth year are available, allowing for cross-validation of responses over time. For analysis, we recoded this continuous variable into a binary variable: where "0" represents individuals aged 18–64, and "1"

represents those aged 65 and older. This classification aligns with the standard definition of older adults as those aged 65 and above [136].

Political affiliation was also measured through the recurring survey administered by YouGov, where participants answered two branching questions: "Generally speaking, do you think of yourself as a…?" with response options including Democrat, Republican, Independent, Other, or Not sure. Those who identified as Democrat or Republican were asked a follow-up question to gauge the strength of partisanship: "Would you call yourself a strong [Democrat/Republican] or a not very strong [Democrat/Republican]?" Those who selected Independent, Other, or Not sure were asked: "Do you think of yourself as closer to the Democratic or the Republican Party?" with response options: the Democratic Party, the Republican Party, Neither, or Not sure. This question wording aligns with well-established items in national surveys such as the American National Election Survey [8]. In our analysis, we recoded political affiliation into a binary variable (Republican vs. non-Republican), as "1" for participants who self-identified as Republican. Participants who did not self-identify as Republican were coded as "0". This recoding further reduces ambiguity and measurement noise associated with midpoints such as "Independent" or "Not sure". It is also important to note that our survey agency YouGov collected these responses repeatedly at 30-day intervals, which accounts for greater stability and reliability over time.

We conducted a validation of this self-reported partisanship measure to assess potential self-report bias. We assessed its convergent validity against both other self-reported items and a behavioral measure of media consumption. The self-reported partisanship aligns with political ideology and voting choices: on a 5-point scale from -2 (very liberal) to 2 (very conservative), Republicans reported being significantly more conservative (M=1.07) than non-Republicans (M=-0.73; t(717) = 26.68, p < .001). Republicans were also significantly more likely to report voting for Donald Trump in 2020 ($\chi^2(1) = 334.28, p < .001$). Crucially, we also validated our self-report measure against a behavioral measure of media consumption using the domain-level partisanship scores developed by Robertson et al. [149]. These scores reflect the partisan lean of a website's audience based on voter file data, with higher scores indicating a more Republican-leaning audience. Our analysis of participants' web browsing data confirms that self-reported Republicans visited domains with a significantly higher pro-Republican bias score (M=0.13) than non-Republicans (M=0.02; t(717) = 152.98, p < .001). These results confirm that our self-reported measure is valid, correlating with conceptually related self-reported and behavioral measures. A full description of our validation is available in Appendix C.

3.4 Data Analysis

To answer our hypotheses and research questions under the user-centric framework, we fit a series of multilevel linear regressions. First, to examine H1—that misinformation compared to hard news, (H1a) is more readable, (H1b) has lower perplexity, (H1c) is lower in sentiment, and (H1d) uses more moral language—we fit the following model:

$$Y_{ij} = \beta_0 + \beta_1 \text{Misinfo}_i + u_j + \epsilon_{ij} \tag{1}$$

This model is at the article-person level, where each observation represents a unique hard news or misinformation article (i) visited by a person (j). We run separate models for each linguistic metric (Y) (i.e., readability, perplexity, sentiment, and moral language). Here, β_0 is the expected value of the linguistic metric for hard news articles, assuming that the error term ϵ has a mean of zero. $Misinfo_{ij}$ is a binary variable representing whether the article is a misinformation article or not, with misinformation articles being coded as "1" and non-misinformation articles (i.e., hard news articles) being coded as "0."

Thus, the intercept term (β_0) represents the expected value of the linguistic metric for hard news articles. β_1 is then the expected difference in the linguistic metric between misinformation and hard news articles, with a positive value representing misinformation articles having a higher linguistic metric and a negative value representing misinformation articles having a lower linguistic metric. u_j captures the random effect on the intercept for person j. We use the p-value associated with β_1 to evaluate H1 using an alpha of .05. Results for these models are found in Table 1. To aid in interpretability, we visualize the predicted values from the model for misinformation and hard news across the linguistic features in Figure 6.

Second, to evaluate RQ1, which examines how linguistic features vary between misinformation and hard news articles across different topics, we fit a model with interaction terms for each topic category. Specifically, we fit the model

$$Y_{ij} = \beta_0 + \beta_1 \text{Misinfo}_i + \boldsymbol{\beta}_{\text{Topic}} \cdot \text{Topic}_i + \boldsymbol{\gamma}_{\text{Interaction}} \cdot (\text{Misinfo}_i \times \text{Topic}_i) + u_j + \epsilon_{ij} \tag{2}$$

This model is also at the article-person level, where each observation represents a unique article (i) classified as either hard news or misinformation, also categorized into one of each of the topics, visited by each person (j). For this analysis, we include binary variables ("1" or "0") representing each topic–Social Issues, U.S. Electoral Politics, COVID-19, and Health–with General News as the reference category. Each article is coded for both its misinformation status (still represented by the binary variable $Misinfo_i$).

In this model, β_0 is the expected value of the linguistic metric for General News articles when they are not misinformation (i.e., hard news). The main effect coefficients in β_{Topic} represent the differences in the linguistic metric for each topic compared to General News (the reference topic) for non-misinformation articles (i.e., hard news). The interaction terms, represented by $\gamma_{Interaction}$, allow us to examine how the effect of misinformation differs across topics. Each γ coefficient in this vector captures the addition effect of misinformation for the specific topic relative to General News. For example, $\gamma_{SocialIssues}$ represents how misinformation status affects the linguistic metric within Social Issue articles compared to General News. By using the p-values associated with each γ coefficient, we assess the statistical significance of the interaction effects, again using an alpha level of .05. The results from these models are found in Table 2. To make these group differences interpretable, we present the predicted values from linguistic feature models in Figure 7.

Finally, for RQ2 and RQ3, we fit a series of multilevel linear models at the person (j)-article (i) level where each observation represents a visit to an article that is nested within people:

$$\begin{aligned} Y_{ij} = \beta_0 &+ \beta_1 \text{Misinfo}_{ij} + \beta_2 \text{OlderAdult}_j + \beta_3 \text{Republican}_j \\ &+ \beta_4 (\text{Misinfo}_{ij} \times \text{OlderAdult}_j) \\ &+ \beta_5 (\text{Misinfo}_{ij} \times \text{Republican}_j) + u_j + \epsilon_{ij} \end{aligned} \tag{3}$$

Y_{ij} represents the outcome of interest (either a linguistic metric or topic model) for a person j visiting article i. The intercept, β_0, captures the baseline level of the outcome. β_1 represents the effect of the article being misinformation on the outcome metric, while β_2 and β_3 represent the effects of age (being an Older adult, coded as "1" if aged 65 or older and "0" if younger than 65), and partisanship (being a Republican, coded as a binary "1" or "0"), respectively. The two-way interaction terms–β_4, β_5–test whether the effect of misinformation on the linguistic feature differs by political affiliation or age group. These two coefficients are used to answer RQ2 and RQ3.

To account for potential individual-level differences and the nested structure of the data, we include a random intercept, u_j, for each person, allowing us to control for unobserved heterogeneity at the person level. The residual error term, ϵ_{ij}, captures unexplained variation within each person-article interaction. Statistical significance is assessed for each coefficient at an alpha level of .05.

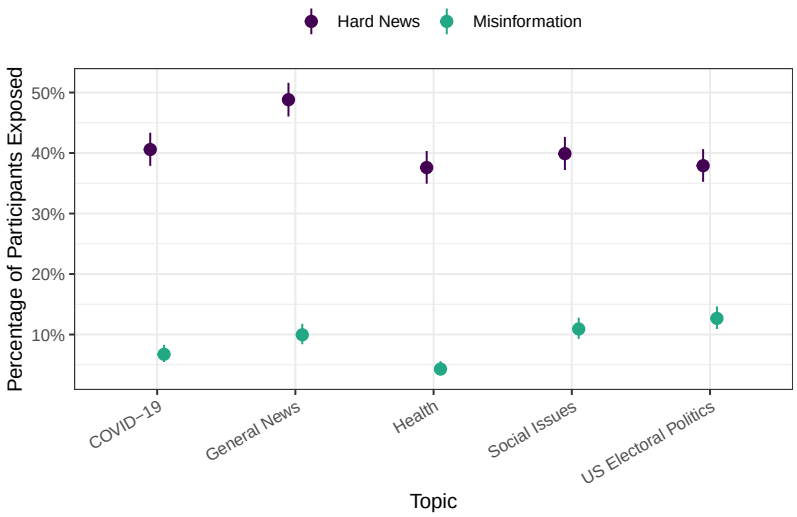


Fig. 5. Descriptive results of topline levels of exposure among the 1,240 participants by misinformation and hard news topics. The x-axis represents different topics. The y-axis is the percentage of participants who were exposed to at least one article of each topic. Points represent point estimates with the lines representing 95% confidence intervals. Purple points are hard news articles. Green points are misinformation articles.

The results for these models are found in Table 3 (topics) and Table 4 (linguistic features). To make the results of these models interpretable, we present predicted values for age groups in Figure 8 and for non-Republicans and Republicans in Figure 9.

4 Results

First, to situate our results in the previous literature that examines topline exposure, we provide descriptive statistics on the percentage of participants who were exposed to at least one misinformation or hard news article of each of the five topical categories (see Figure 5). We found that the most consumed type of misinformation articles was about U.S. Electoral Politics, with 12.7% (95% CI [10.9, 14.7]) of participants being exposed to at least one such article. The next topic with the highest level of exposure was misinformation about social issues with 10.9% (95% CI [9.3, 12.8]) of participants exposed. The least exposed types were COVID-19 (6.7% exposed, 95% CI [5.4, 8.3]) and Health (4.3% exposed, 95% CI [3.3, 5.6]).

4.1 Linguistic Features of Misinformation vs. Hard News

For our first set of hypotheses (H1a-c), we conducted a series of mixed-effects regression analyses to examine the linguistic style differences between misinformation and hard news across the four key linguistic features: readability, sentiment, morality, and perplexity. The statistical results are presented in Table 1 and visualized in Figure 6. The intercept coefficients (β_0) represented the mean values for hard news, while the main effect of misinformation (β_1) indicated the mean differences for misinformation articles compared to hard news. Since lower readability scores indicate easier-to-read content, the negative coefficient for misinformation in Table 1 suggests that misinformation articles were significantly more readable than hard news (β_1 = -1.17, SE = 0.28, p < 0.001). Misinformation articles were also associated with significantly more negative sentiment (β_1 = -1.61, SE = 0.09, p < 0.001), while hard news (β_0 = 0.96, SE = 0.11, p < 0.001) had significantly

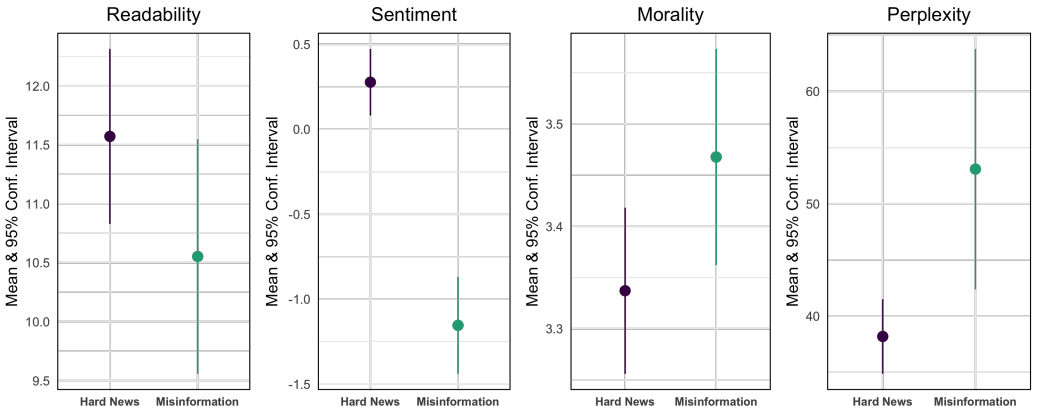


Fig. 6. Linguistic metrics (readability, sentiment, morality, and perplexity) by hard news and misinformation. Each panel represents a different linguistic feature. The x-axis is whether the articles are from hard news or misinformation domains. The y-axis represent the given linguistic feature's measurement. Each point represents the point estimates of the predicted probabilities from the respective models (Equation 1) with the lines representing 95% confidence intervals.

Table 1. Linear Mixed Effects Models of Linguistic Style Differences Between Misinformation and Hard News.

	Readability			Sentiment			Morality			Perplexity		
Predictors	Estimates	SE	*p*	Estimates	SE	*p*	Estimates	SE	*p*	Estimates	SE	*p*
β_0 Intercept	11.65	0.38	<0.001	0.96	0.11	<0.001	2.99	0.04	<0.001	41.51	1.62	<0.001
β_1 Misinfo	-1.17	0.28	<0.001	-1.61	0.09	<0.001	0.56	0.03	<0.001	-1.37	4.20	0.744
Random Effects												
σ^2	281.87			29.47			3.26			69860.85		
τ_{00} (caseid)	79.71			6.25			1.13			566.98		
ICC	0.22			0.17			0.26			0.01		
N	719 (caseid)			719 (caseid)			719 (caseid)			719 (caseid)		
Observations	150905			150904			150905			150905		
Marginal R^2 / Conditional R^2	0.000 / 0.221			0.003 / 0.178			0.003 / 0.259			0.000 / 0.008		

higher frequency of moral term terms compared to hard news (β_1 = 0.56, SE = 0.03, p < 0.001). However, we found no significant difference between misinformation and hard news in perplexity (β_1 = 1.37, SE = 4.2, p = 0.744). Therefore, the results support H1a, H1c, and H1d, while do not find support for H1b.

4.2 Linguistic Styles Across Topics

Our first research question acknowledged that contextual factors, such as the topic, can influence linguistic styles. For RQ1, we compared the linguistic features of misinformation and hard news across five topics: social issues, COVID-19, health, U.S. electoral politics, and general news, with the latter serving as the reference category. The results of the mixed effect regression analysis are presented in Table 2 and visualized in Figure 7.

While we found support for H1a that misinformation is generally more readable than hard news, a closer examination of topic differences revealed that this difference in readability is primarily driven by the topic of U.S. politics. Overall, there were minimal differences in readability between misinformation and hard news, except for U.S. electoral politics. Specifically, hard news on U.S. electoral politics (β = 0.87, SE = 0.14, p < 0.001) was significantly harder to read than misinformation. It was the only topic where misinformation was more difficult to read than hard news.

Table 2. Linear Mixed Effects Models of Linguistic Style Differences Between Misinformation and Hard News Across Topics

	Readability			Sentiment			Morality			Perplexity		
Predictors	Estimates	SE	p	Estimates	SE	p	Estimates	SE	p	Estimates	SE	p
β_0 Intercept	11.61	0.38	<0.001	2.56	0.10	<0.001	2.26	0.04	<0.001	48.15	1.91	<0.001
β_1 Misinfo	-1.33	0.76	0.081	-0.29	0.24	0.224	1.00	0.08	<0.001	76.60	11.91	<0.001
β_{COVID19}	-0.71	0.18	<0.001	-2.59	0.05	<0.001	1.25	0.02	<0.001	-10.72	2.72	<0.001
β_{Health}	0.01	0.24	0.952	-2.74	0.07	<0.001	1.59	0.02	<0.001	-20.81	3.63	<0.001
$\beta_{\text{SocialIssues}}$	-0.36	0.16	0.026	-4.73	0.05	<0.001	1.92	0.02	<0.001	-7.54	2.50	0.003
$\beta_{\text{USelecPolitics}}$	0.87	0.14	<0.001	-1.37	0.04	<0.001	0.65	0.01	<0.001	-10.74	2.17	<0.001
$\gamma_{\text{COVID19}\times\text{Misinfo}}$	1.34	0.97	0.166	-1.55	0.30	<0.001	-0.61	0.10	<0.001	-81.08	15.17	<0.001
$\gamma_{\text{Health}\times\text{Misinfo}}$	0.17	1.40	0.906	-3.26	0.44	<0.001	-2.34	0.14	<0.001	-54.87	21.88	0.012
$\gamma_{\text{SocialIssues}\times\text{Misinfo}}$	0.90	0.90	0.318	-0.08	0.28	0.779	-0.69	0.09	<0.001	-81.52	14.15	<0.001
$\gamma_{\text{USelecPolitics}\times\text{Misinfo}}$	-0.86	0.85	0.312	-0.83	0.27	0.002	-0.71	0.09	<0.001	-91.08	13.26	<0.001
Random Effects												
σ^2	281.70			27.64			2.92			69815.29		
τ_{00} (caseid)	79.64			5.22			0.97			551.09		
ICC	0.22			0.16			0.25			0.01		
N	719 (caseid)			719 (caseid)			719 (caseid)			719 (caseid)		
Observations	150905			150904			150905			150905		
Marginal R^2 / Conditional R^2	0.001 / 0.221			0.079 / 0.225			0.115 / 0.336			0.001 / 0.009		

In terms of sentiment, we observed considerable variations across topics. In general, misinformation articles involved more negative sentiment than general news. However, the sentiment tone between misinformation and hard news varied across topics. Hard news overall had a more positive tone for general news and politics, a neutral tone for COVID-19 and health, and a negative tone for social issues. In contrast, misinformation tended to exhibit a consistently negative tone, except for general news. The interaction between each topic and misinformation, as shown in Table 2, revealed that misinformation articles on COVID-19 (β = -1.55, SE = 0.30, p < 0.001), health (β = -3.26, SE = 0.44, p < 0.001), and U.S. politics (β = -0.83, SE = 0.27, p = 0.002) demonstrated more negative sentiment than hard news, with COVID-19 and health misinformation showing the most pronounced negativity.

Regarding morality, in general, misinformation tended to incorporate more moral language. Still, topic-specific analysis revealed a nuanced pattern: for health, hard news used a significantly higher level of moral terms than misinformation. Conversely, hard news on COVID-19, health, social issues, and U.S. politics all used more moral language compared to misinformation, with social issues showing the highest usage of moral terms.

Finally, while we previously did not observe a difference in perplexity between misinformation and hard news overall (H1b), a topic-specific analysis did reveals topic-related differences. Misinformation related to general news was more perplexing than hard news, with health topics showing a similar trend. In contrast, for other topics, the perplexity scores between hard news and misinformation were comparable (e.g., COVID-19 and social issues). In contrast, political news involved lower perplexity for misinformation than hard news.

4.3 Older Adults and Republicans' Experience of Misinformation

To examine how the misinformation that older adults consume is linguistically different compared to younger groups, we estimated a series of linear mixed-effect models. First, we focused on topical differences, with results presented in Table 3. The findings showed that the consumption of hard news among older adults did not significantly differ from that of their younger counterparts, except U.S. electoral politics (β_3 = 0.04, SE = 0.02, p = 0.027). Older adults consumed significantly more election news than the younger group. Our hypothesis (H2) predicted that older adults would be more likely to consume misinformation in health, COVID-19 and electoral politics than younger

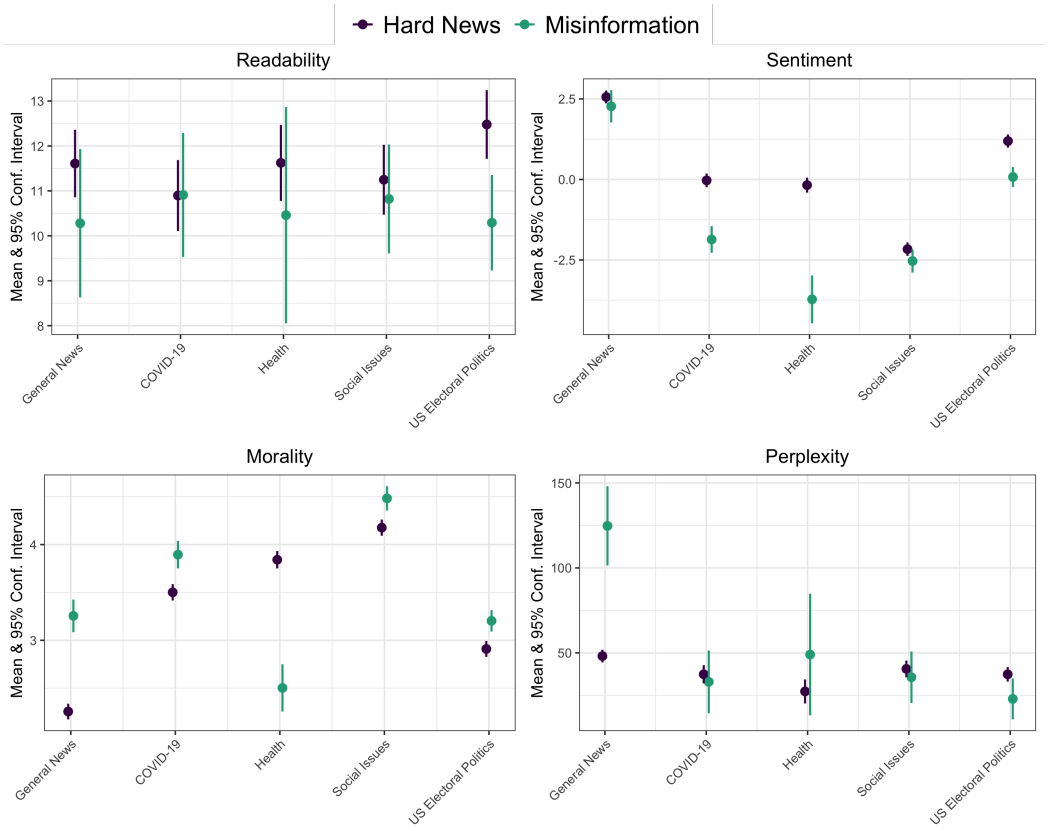


Fig. 7. Linguistic metrics (readability, sentiment, morality, and perplexity) by hard news and misinformation. Each panel represents a different linguistic feature. The x-axis represent different topics of articles. The y-axis represent the measurement of the given linguistic feature. Each point represents the point estimates of the predicted values from the respective models (Equation 2) with the lines representing 95% confidence intervals. Purple points represent articles from hard news domains. Green points represent articles from misinformation domains.

adults. We found that older adults consumed significantly more misinformation related to COVID-19 (β_5 = 0.04, SE = 0.01, p < 0.001) and health (β_5 = 0.02, SE = 0.01, p = 0.025) but not to electoral politics, partially supporting H2.

In terms of linguistic differences (see Table 4), our hypothesis predicted that relative to non-older adults, older adults would consume misinformation involving (a) positive sentiment, (b) lower readability, and (c) lower perplexity. The results were inconsistent with this hypothesis. Overall, older adults consumed news with more negative sentiment (β_3 = -0.63, SE = 0.23, p = 0.007) than younger adults. This gap in negative sentiment between younger and older adults was significantly larger for misinformation, with older adults consuming significantly more negative sentiment misinformation (β_5 = -1.99, SE = 0.19, p < 0.001), as can be seen in Figure 8. Furthermore, we found no main effect of age on readability (β_5 = -0.67, SE = 0.83, p = 0.419) and perplexity (β_5 = 3.07, SE = 3.54, p = 0.386). There was also no significant interaction effect between misinformation and older adults, meaning that for misinformation, the levels of readability (β_5 = 0.19, SE = 0.60, p =

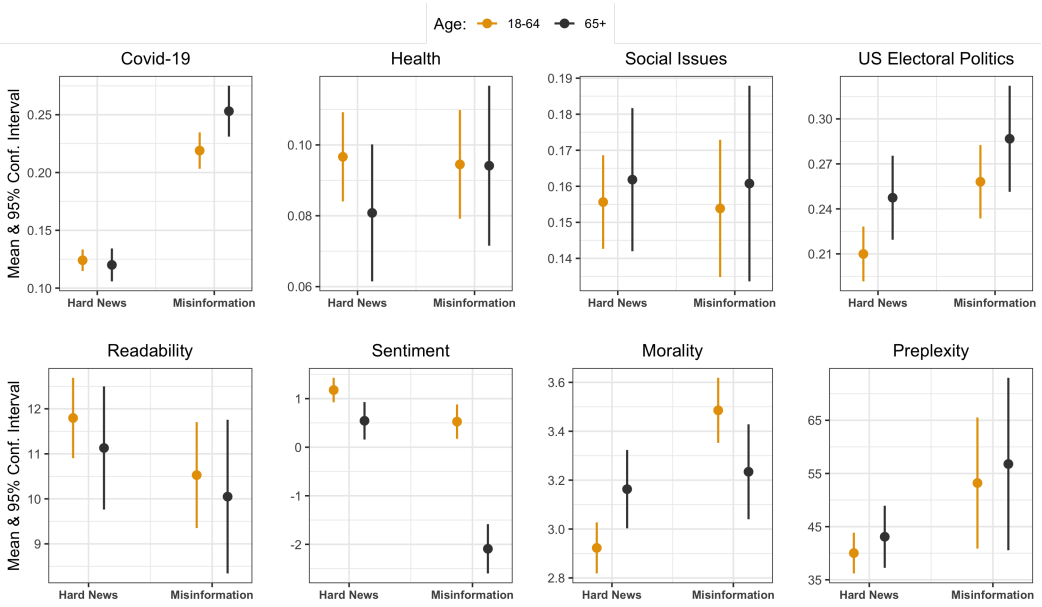


Fig. 8. Visualization for age across topics and linguistic patterns. Note: Topics (COVID-19, Health, Social Issues, and U.S. Electoral Politics) and linguistic metrics (readability, sentiment, morality, and perplexity) by hard news and misinformation by age group. Each panel represents a different topical or linguistic model (Equation 3). The x-axis represent articles from either hard news or misinformation domains. The y-axis represent the predicted values from the models. Each point represents the point estimates of the predicted values from the respective models (Equation 3) with the lines representing 95% confidence intervals. Yellow points represent users aged 18-64. Black points represent users aged 65+.

Table 3. Linear Mixed-Effects Model Analysis of Individual Differences in Misinformation Topics.

	General News			COVID-19			Health			Social Issues			US Electoral Politics		
Predictors	Estimates	std. Error	*p*	Estimates	SE	*p*	Estimates	SE	*p*	Estimates	SE	*p*	Estimates	SE	*p*
β_0 Intercept	0.40	0.01	<0.001	0.13	0.01	<0.001	0.10	0.01	<0.001	0.16	0.01	<0.001	0.22	0.01	<0.001
β_1 Misinfo	-0.09	0.01	<0.001	0.12	0.01	<0.001	0.00	0.01	0.658	-0.03	0.01	0.001	0.00	0.01	0.974
β_2 OlderAdult	-0.03	0.02	0.160	-0.00	0.01	0.641	-0.02	0.01	0.174	0.01	0.01	0.604	0.04	0.02	0.027
β_3 Republican	0.08	0.02	<0.001	-0.02	0.01	0.041	-0.01	0.01	0.595	-0.01	0.01	0.595	-0.05	0.02	0.001
β_4 Misinfo ×*OlderAdult*	-0.05	0.01	0.001	0.04	0.01	<0.001	0.02	0.01	0.025	0.00	0.01	0.950	-0.01	0.01	0.487
β_5 Misinfo ×*Republican*	-0.20	0.01	<0.001	-0.12	0.01	<0.001	-0.02	0.01	0.001	0.13	0.01	<0.001	0.22	0.01	<0.001
Random Effects															
σ^2	0.14			0.08			0.04			0.09			0.13		
τ_{00} (caseid)	0.05			0.01			0.02			0.02			0.03		
ICC	0.25			0.09			0.31			0.14			0.21		
N	719 (caseid)			719 (caseid)			719 (caseid)			719 (caseid)			719 (caseid)		
Observations	150905			150905			150905			150905			150905		
Marginal R^2 / Conditional R^2	0.017 / 0.266			0.004 / 0.090			0.001 / 0.315			0.002 / 0.146			0.010 / 0.216		

0.754) and perplexity (β_5 = 0.51, SE = 9.04, p = 0.955) consumed by older adults were not statistically different from those of younger groups.

Unexpectedly, older adults consumed news that included more moral language (β_3 = 0.24, SE = 0.10, p = 0.013) than younger adults. However, for misinformation, older adults showed a preference for content that uses fewer moral terms (β_5 = -0.49, SE = 0.06, p < 0.001).

Finally, to address RQ2a, we examined how Republicans consume misinformation differently from non-Republicans. We first examined the topic. As shown in Table 3, we found Republicans consumed more content about general news and electoral politics, but less COVID-19 news than

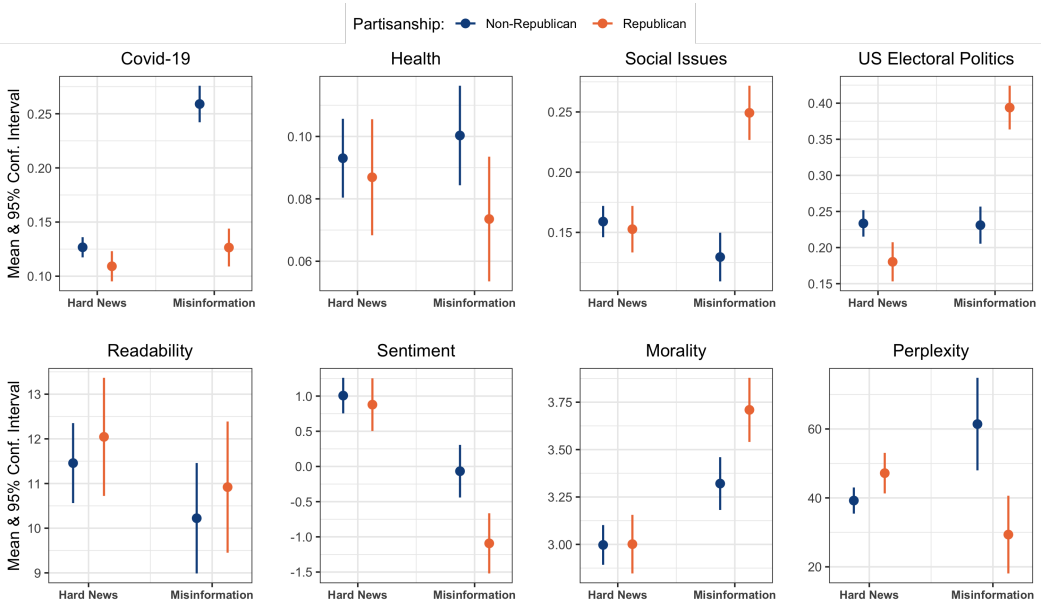


Fig. 9. Visualization for partisanship across topics and linguistic patterns. Note: Topics (COVID-19, Health, Social Issues, and U.S. Electoral Politics) and linguistic metrics (readability, sentiment, morality, and perplexity) by hard news and misinformation by partisanship. Each panel represents a different topical or linguistic model (Equation 3). The x-axis represents articles from either hard news or misinformation domains. The y-axis represents the predicted values from the models. Each point represents the point estimates of the predicted values from the respective models (Equation 3) with the lines representing 95% confidence intervals. Navy points represent non-Republican users. Orange points represent Republican users.

non-Republicans. Relative to non-Republicans, Republicans consumed less misinformation on COVID-19 (β_4 = -0.12 , SE = 0.01, p < 0.001) and health (β_4 = -0.02 , SE = 0.01, p = 0.001), but more on social issues (β_4 = 0.13 , SE = 0.01, p < 0.001) and electoral politics (β_4 = 0.22 , SE = 0.01, p < 0.001).

To address RQ2b, we next examined the sentiment consumed by Republicans compared to non-Republicans. The significant interaction term between misinformation and partisanship in Table 4 indicated that Republicans engaged with misinformation characterized by significantly more negative sentiment (β_4 = -0.90 , SE = 0.18, p < 0.001), and inclusion of more moral language β_4 = 0.38 , SE = 0.06, p < 0.001), and lower perplexity (β_4 = -40.01, SE = 8.64, p < 0.001) compared to non-Republicans. These results were exclusively for misinformation, as we observed no significant differences in linguistic features among partisans regarding hard news. Perplexity (β_2 = 7.95, SE = 3.58, p = 0.026) was an exception, as Republicans preferred hard news with greater lexical diversity compared to non-Republicans.

5 Discussion

5.1 Summary

Our study revealed several key findings that underscore the importance of a user-centric approach in understanding misinformation consumption. First, consistent with prior research [28], we found that misinformation consumed by users was generally easier to read, exhibited higher negative

Table 4. Linear Mixed-Effects Model Analysis of Individual Differences in Misinformation Linguistic Styles.

	Readability			Sentiment			Morality			Perplexity		
Predictors	Estimates	SE	p	Estimates	SE	p	Estimates	SE	p	Estimates	SE	p
β_0 Intercept	11.67	0.51	<0.001	1.21	0.14	<0.001	2.92	0.06	<0.001	38.27	2.18	<0.001
β_1 Misinfo	-1.29	0.48	0.007	-0.45	0.16	0.004	0.48	0.05	<0.001	22.03	7.36	0.003
β_2 OlderAdult	-0.67	0.83	0.419	-0.63	0.23	0.007	0.24	0.10	0.013	3.07	3.54	0.386
β_3 Republican	0.59	0.82	0.471	-0.13	0.23	0.577	0.00	0.10	0.966	7.95	3.58	0.026
β_4 Misinfo ×*OlderAdult*	0.19	0.60	0.754	-1.99	0.19	<0.001	-0.49	0.06	<0.001	0.51	9.04	0.955
β_5 Misinfo ×*Republican*	0.11	0.57	0.848	-0.90	0.18	<0.001	0.38	0.06	<0.001	-40.01	8.64	<0.001
Random Effects												
σ^2	281.87			29.44			3.25			69849.96		
τ_{00} (caseid)	79.81			6.10			1.12			572.19		
ICC	0.22			0.17			0.26			0.01		
N	719 (caseid)			719 (caseid)			719 (caseid)			719 (caseid)		
Observations	150905			150904			150905			150905		
Marginal R^2 / Conditional R^2	0.001 / 0.221			0.009 / 0.179			0.006 / 0.261			0.000 / 0.008		

sentiment, and employed more moral language than hard news. Contrary to previous findings, misinformation did not display lower perplexity overall than hard news.

Second, our results highlighted the critical role of contextual factors, such as news topics, in shaping the linguistic features of misinformation [98]. We found that the increased readability of misinformation was primarily driven by content related to U.S. electoral politics and was not consistent across other topics. Misinformation consistently adopted a more negative tone, especially in COVID-19 and health topics, contrasting with the more neutral tones in hard news. Health misinformation contained far less moral language than hard news about health, while for every other topic misinformation contained more moral language. While overall differences in perplexity were not observed, significant variations emerged when evaluating specific topics.

Third, we observed heterogeneity in misinformation consumption among groups that are frequently found to be more exposed to misinformation, specifically, older adults and Republicans (e.g., [107]). Older adults consumed significantly more misinformation about COVID-19 and health issues than younger adults, partially supporting our second hypothesis. Contrary to our hypothesis grounded in assumed cognitive preferences, older adults consumed misinformation with more negative sentiment and fewer moral terms, and we found no significant differences in readability or perplexity.

Finally, we found that Republicans engaged with misinformation characterized by more negative sentiment, greater use of moral language, and lower perplexity compared to content consumed by non-Republicans. They also consumed misinformation on different topics, including less about COVID-19 and health issues but more about social issues and electoral politics.

These findings emphasize that misinformation is not a monolithic phenomenon but varies significantly across topics and user characteristics. By focusing on the actual content consumed by individuals, we show important variations in linguistic features and topic preferences that may be overlooked in studies that do not adopt a user-centric approach to misinformation. This has important implications for designing targeted interventions, such as tailoring strategies to address the specific linguistic patterns and topics relevant to different user groups, which could enhance the effectiveness of efforts to combat misinformation [82], particularly considering that current platform-level approaches may have limited effectiveness [34].

5.2 Theoretical Implications

5.2.1 Importance of a user-centric perspective. Our study demonstrates the importance of taking a user-centric approach to examine the content that people actually consumed. Examining consumed misinformation is important because it reveals the types of misinformation that may be genuinely

resonating with users, highlighting the psychological, cognitive, and demographic factors that drive misinformation engagement. While past studies have focused on quantifying the supply-side of misinformation [28, 81, 143, 167], a user-centric approach can speak more to the demand side of misinformation. Examining where supply and demand may differ provides a clearer understanding of the real-world impact of misinformation and provides insights into how it may spread.

In addition, this approach allows our study to be in conversation with extant web-browsing data misinformation exposure work. This past work almost exclusively operates at the domain level and does not consider the actual content of the webpages people visit. This past work has indeed contributed significantly to our understanding of misinformation and news exposure dynamics, e.g., [92, 93, 180, 191]. Thus, our study, examining the content of the pages, extends this body of knowledge into a more granular direction. For example, work that operates at the domain level can speak to topline exposure to categories of websites, but more granular categories may have much lower exposure, e.g., see Figure 5.

5.2.2 Linguistic differences in misinformation. Our study largely replicates Carrasco-Farré [28]'s study on the linguistic characteristics of misinformation, replicating with our user-centric approach that misinformation tends to be more readable, more negative, and contains more moral language. We failed, however, to replicate the finding of lower levels of perplexity in misinformation. One possible explanation for this discrepancy may lie in the methodology used to measure perplexity in this study. While Carrasco-Farré [28] used a corpus of only misinformation news and hard news to train their perplexity model, our study utilized the much larger language corpus that was used to train GPT-2. This database reflects a much broader range of content from the Internet than prior work. Therefore, our perplexity results should be interpreted with respect to how much perplexity was observed from language on the Internet. With this approach, we saw no difference in perplexity across misinformation and hard news.

The initial argument suggesting that misinformation may exhibit lower perplexity posits that texts with lower perplexity require less cognitive effort, making them more likely to be believed [28]. However, from a user-centric perspective, highly perplexed language does not necessarily require more cognitive processing. In fact, studies have shown that participants engage in less cognitive processing when reading more complex misinformation articles [91]. As a result, individuals may be more likely to believe misinformation presented in a complex way. This explanation is consistent with the deception detection literature, which also suggested that complex messages may be considered more truthful than simpler messages [119, 134]. Finally, we also observed an important influence of the topic. At an aggregate level, without accounting for topic specificity, it may be challenging to identify differences in perplexity.

5.2.3 Contextual heterogeneity and the implication for language and deception. Our findings of substantial linguistic differences in misinformation across news topics are consistent with the Contextual Organization of Language and Deception (CoLD) framework by Markowitz and Hancock [98], which emphasizes the crucial role of context in shaping deceptive communication. The linguistic variations we observed, particularly the increased negative sentiment and moral language in health-related misinformation, highlight the importance of topic as a contextual factor influencing language use. The CoLD framework suggests that deceptive language is not monolithic but adapts to specific contexts to enhance credibility and persuasion. By demonstrating that misinformation extends beyond political content and varies linguistically depending on the topic, our study extends the interpersonal deception insights of the CoLD framework into the broader field of misinformation. This alignment suggests that understanding the contextual nuances of misinformation is essential for developing effective interventions, as it highlights the need to tailor strategies according to the unique linguistic patterns associated with different topics, thus building on past work that

demonstrates how the effectiveness of misinformation corrections depends critically on issue type and source credibility [45].

5.2.4 Older adults and the language of misinformation. The finding that older adults consume more news with negative sentiment appears counterintuitive when viewed through socioemotional selectivity theory (SST). SST posits that as individuals age and perceive their life horizon shortening, they prioritize emotionally meaningful goals, typically favoring positive over negative information to enhance emotional well-being[29, 147]. However, our results indicate that older adults consumed more negative misinformation, particularly related to health and COVID-19 issues.

One possible explanation for this discrepancy is that older adults encounter more misinformation overall, which is generally negative in tone. Misinformation tends to attract greater attention and engagement due to its emotionally charged and often alarming content [15, 35, 102, 176, 183]. Within this framework, our findings align with the dual-process theories of cognition [47, 170], which suggest that emotionally charged content is more likely to activate intuitive thinking and rely on heuristics and emotional shortcuts. When exposed to such content, individuals, particularly older adults with reduced cognitive capacity [26, 182] and lower digital literacy [24], may be more susceptible to believing and sharing such content without engaging in careful and analytical evaluation of the content's accuracy. These findings imply that misinformation purveyors strategically design content to exploit older adults' digital literacy and cognitive vulnerabilities.

Another explanation is that topic relevance, especially health-related, may outweigh the general positivity bias described by SST. Given that older adults prioritize health-related information due to its personal relevance [21], they may inadvertently consume more negative content within this domain. The heightened focus on health and COVID-19 topics—areas where misinformation is both prevalent and particularly negative—could override their general preference for positive information.

Past research found that older adults are more vulnerable to misinformation due to higher exposure rates [64, 109] and frequent targeting [24]. However, some studies indicate that older adults may also demonstrate better discrimination abilities when evaluating news content [163]. Our findings contribute to this debate by highlighting topic-specific differences in misinformation engagement among older adults, suggesting that susceptibility may not be uniform across different types of misinformation.

Our findings have significant implications for understanding the vulnerability of older adults to health-related misinformation. Continuous exposure to negative news about COVID-19 and health issues can have detrimental effects on mental health, potentially leading to increased stress, anxiety, or depression. If this distressing content is false or misleading, it could also adversely affect physical health by influencing harmful behaviors or discouraging beneficial ones, such as adhering to medical advice or public health guidelines.

Given the significantly higher exposure of older adults to negative health and COVID-19 misinformation, future research should focus on examining the effects of this exposure on their mental and physical health over time. There is a critical need to develop effective interventions tailored to older adults that reduce their exposure to misinformation and enhance their ability to discern credible health information. This highlights the need for interventions that go beyond factual correction and instead focus on inoculating users against emotional manipulation. Effective strategies include digital literacy programs that enhance critical thinking skills, friction-based interventions that prompt users to pause and reflect before sharing [12, 49], and targeted messaging that raises awareness of emotionally manipulation tactics [94].

5.2.5 Partisan differences in topic preference and linguistic patterns. We observe differences in the topics of misinformation consumed by Republicans and non-Republicans. Republicans consumed

less election-related hard news but more election-related misinformation. This observation may be due to the changing information ecosystem [17, 146, 173], characterized by the rise of social media and partisan niche media [33, 41, 158], as well as political attacks on established institutions, e.g., [17], such as mainstream news [27, 43, 103]. Therefore, hard news websites (i.e., "the mainstream news media") may be perceived by Republicans as not representing their interests [104, 156]. This distrust is likely driving them to eschew hard news websites and instead to seek information from alternative websites, particularly around elections [69, 142].

The findings that non-Republicans are more likely to consume misinformation about health and COVID-19 may be surprising to some, particularly because COVID-19 skepticism and vaccine hesitancy in the United States is generally found among conservatives [71, 78, 112, 144] and conservative media [161]. However, past work has found that sometimes when people encounter health misinformation, they think the claim is so outlandish they want to investigate it further [95]. Another possibility is that non-Republicans are more interested in health and COVID-19 information generally, as evidenced by them being more likely to visit hard news about health and COVID-19 (see Table 3). Thus, in their search for information on these topics, they may be incidentally exposed to misinformation on these topics [20, 108]. These potential explanations are bolstered by the findings that show that exposure to misinformation does not necessarily equate to belief in that misinformation [166].

We also found evidence of asymmetrical consumption in linguistic styles, namely that Republicans tended to consume less perplexing misinformation that contains more moral terms and negative sentiments than non-Republicans. These findings are consistent with meta-analyses and a growing body of literature documenting partisan asymmetry in misinformation consumption and detection, showing that Republicans, compared to Democrats, exhibit a reduced ability to distinguish true from false news and a slightly more pronounced true-news bias [163]. This pattern complements recent findings that exposure to partisan content can shape affective polarization [137], even while evidence suggests that mere exposure to like-minded sources alone does not directly increase polarization [125]. Together, these findings suggest that the specific linguistic features we identify—rather than just source identity—may play a crucial mediating role in how content affects beliefs and attitudes, creating a potential bidirectional relationship between the linguistically distinctive content Republicans consume and their attitude formation.

One plausible explanation for this asymmetrical consumption in linguistic styles is that different political groups are exposed to distinct informational ecosystems characterized by echo chambers and micro-targeting. For example, studies indicate that Republican politicians use more belief-based statements over evidence-based ones [80], and misinformation tends to favor Republican positions [54]. Additionally, Republicans are exposed to more misinformation overall [107]. Another explanation lies in the between-party asymmetry in motivational and cognitive orientations [39, 55, 75, 133], as Republicans may engage more in intuitive reasoning.

5.2.6 LLM hybrid pipeline. Finally, our novel LLM-enhanced topic modeling approach, integrating unsupervised topic modeling, human-in-the-loop validation, and LLM insights, maximizes the strengths of each method while mitigating their limitations. Unsupervised topic modeling traditionally requires extensive human validation, while zero-shot LLMs often lack transparency and replicability. Our framework begins with the transparent establishment of topics, applying these labels across a larger text corpus through LLMs to achieve both interpretability and efficiency. Future research can adopt this framework to investigate various topics and concepts. Furthermore, given LLMs' capabilities in handling multilingual corpora [144], this framework can be expanded to diverse linguistic contexts or cross-lingual, cross-platform research.

5.3 Implications for Design

Our study adopts a user-centric perspective, demonstrating that the linguistic features of hard news and misinformation vary according to topical patterns and that these variations appeal to different audience segments. These insights have implications for platform design, digital literacy interventions, and future misinformation research.

5.3.1 Leveraging Linguistic Cues in Content Ranking. One potential application of our findings is customizing content ranking algorithms to reduce exposure to misinformation, particularly for vulnerable populations. Our study aligns with previous research showing that misinformation often uses emotional manipulation and appeals to outrage [17, 102, 175]. It is also consistent with machine learning studies demonstrating that linguistic features such as readability and lexical diversity can reliably classify misinformation, e.g., [81, 143, 167]. Together, these findings suggest that platforms could customize algorithms to down-rank content that may disproportionately target or mislead certain demographic groups. For instance, health-related misinformation with highly negative emotional tones from unreliable sources could be flagged for additional scrutiny before being widely distributed.

5.3.2 Expanding Digital Media Literacy Interventions. Beyond platform-based designs, our findings point to the need for targeted digital literacy programs that take into account differences in susceptibility to misinformation among demographic groups. Our observation of negative sentiment in misinformation aligns with pre-bunking and inoculation research [87, 150, 151, 175], which highlights the importance of teaching users to recognize linguistic cues and rhetorical techniques such as fear-mongering and emotionally charged language. The distinct linguistic patterns we observe across different audience segments suggest that, similar to how relationship-building approaches have improved engagement with legitimate news sources [162], misinformation interventions should be tailored to specific audience preferences and consumption patterns. Furthermore, our findings suggest that pre-bunking interventions should also address the increased use of moral language in misinformation, particularly when targeting Republicans and older adults.

5.3.3 The Challenges of a Purely Linguistic Approach. Although linguistic cues provide valuable insights, we also observe critical challenges in relying solely on this approach, as we have already found that linguistic patterns vary across topics. Additionally, misinformation purveyors can adapt their language over time to evade detection, and linguistic patterns may exhibit temporal variations. Future studies should combine linguistic analysis with human detection to develop more robust misinformation interventions by developing customized strategies that encourage user engagement in accuracy assessment. While linguistic analysis provides valuable insights, complementary strategies might include targeting content creators through incentives and nudges, with recent evidence showing that even gentle interventions can effectively reduce misinformation sharing by political elites [10]. This source-level approach could be particularly effective in addressing misinformation, given the unique linguistic patterns we've identified across different demographic groups. Future studies should investigate the effectiveness of context-aware accuracy nudges that prompt users to reflect on accuracy before sharing, particularly when content contains linguistic signals indicative of misinformation. For example, since older adults tend to consume more negative and health-related misinformation, it is possible to prompt them to think about accuracy on news that has highly emotional framing in health misinformation [134], and provide them with tips to engage in verification to help them discern misinformation from true news [63, 168].

5.3.4 Further Extending User-Centric Approach. A user-centric perspective is crucial for understanding how individuals consume misinformation content across diverse demographic groups.

Future research should explore additional individual differences, such as cognitive abilities, digital media literacy, and multilingual backgrounds, and examine how topical and linguistic cues appeal to these populations to inform tailored interventions. The partisan differences we observe in linguistic features may be further amplified by algorithmic systems that preferentially expose users to ideologically aligned content while reducing exposure to opposing viewpoints [13, 22, 25, 50, 62, 86, 116, 121, 140, 189], potentially creating a reinforcing cycle between the linguistic patterns we identify and users' information diets. More work is also needed to examine the underlying mechanisms that explain why certain individuals are more susceptible to specific linguistic markers of misinformation.

In addition, while our study identifies key linguistic markers of misinformation (e.g., sentiment, moral language, readability), future research should expand this scope by examining a more nuanced set of linguistic cues, tracking how these patterns evolve over time, and building dynamic detection models that adapt to shifting styles. Our study implements an LLM-enhanced approach for topic classification; other studies in the field also leverage LLMs for misinformation detection [89, 177]. Future studies should leverage LLMs to analyze topical and linguistic cues and develop real-time, personalized interventions tailored to user characteristics.

5.4 Limitations and Future Research

This paper has several important limitations. First, despite the extensive computational analyses in this research, this paper's exclusive focus on textual content and misinformation overlooks other media included in the content, particularly visual misinformation. Visual content—such as images, illustrations, and videos—plays an increasingly significant role in disseminating false information [67, 77, 79, 101, 117, 128, 152, 154, 179]. Research also highlights the superior effects of visuals in triggering stronger emotional responses and greater online engagement than their textual counterparts [30, 164], underscoring the importance of examining visual misinformation in future studies. Future research can integrate web tracking, content scraping, and computer vision methods to capture the presence of visual and multimodal misinformation, analyze their distinctive content features, and assess their roles in the spread of misinformation online and design interventions [138, 187].

Next, like other studies in the misinformation detection literature, our work relies on linguistic features such as sentiment and readability [28, 81, 143, 167], which have limitations in fully capturing the complexities of misinformation. Our study takes an initial step in addressing this limitation by examining how linguistic characteristics vary across topics; however, we recognize that challenges remain. Specifically, (1) we included only a limited set of four linguistic features, (2) these features can evolve over time and be deliberately manipulated by misinformation spreaders, and (3) linguistic patterns likely interact with other misinformation signals, such as network structure and source credibility, in shaping misinformation dynamics. Future research should examine additional linguistic features [3, 81, 167], such as lexical terms and information language, discrete sentiments, and integrate additional signals of misinformation, including engagement patterns, network characteristics, and credibility assessments, to develop a more comprehensive understanding of misinformation in a user-centric approach.

In addition, we build upon past work that examines exposure to misinformation and news at the domain level by scraping the content of misinformation and hard news pages. However, the limitation of defining misinformation and hard news websites at the domain level remains a limitation shared by most web-tracking studies. This limitation persists in the field because of the difficulty in manually labeling articles. The process of producing original, high-quality fact checks is extensive, e.g., [48]. Relying on existing fact checks to cross-reference is also fraught with inconsistencies, both when examining what gets fact-checked and also the ratings on specific

claims [83]. Automated methods for identifying misinformation have also been attempted, but Conti et al. [36] found that classification F-scores never exceeded 0.7, a minimum standard for accurate classification. Wisdom of the Crowds methods have shown promising results [5, 74, 99], and future work should seek to integrate Wisdom of the Crowds with a user-centric approach that examines the information that people are actually engaging with online, e.g., [111]. Furthermore, our results rely on self-reported measures for individuals' partisanship and age. These self-reports may induce bias. We examine this potential bias and find that our measures are correlated with other self-reported measures and robust to using behavioral measures of partisan media consumption instead of self-reported partisanship. For more details, please see Appendix C.

Furthermore, our analysis examined age and partisanship as separate demographic factors. A limitation of this approach is that it may obscure important differences that exist at the intersections of these identities. Future work should adopt a more intersectional lens for understanding how overlapping social identities shape engagement within socio-technical systems. For instance, it is possible that the media diets and online experiences of a younger Republican, shaped in digitally-native partisan communities, differ significantly from those of an older Republican accustomed to traditional broadcast media. This could lead to divergent patterns of susceptibility, where one group is more vulnerable to meme-based, low-perplexity content, while the other is more influenced by misinformation that mimics the authoritative tone of traditional news. Similarly, the topical entry points for misinformation may differ; older Republicans might be uniquely targeted by content that bridges health anxieties with political conspiracies. By exploring these demographic intersections, future research can develop a more fine-grained understanding of misinformation vulnerability and move toward interventions that are not only tailored to single demographic groups, but to the specific experiences of intersectional identities.

This study only examines data from a panel of American adults. Misinformation is a global challenge [90]. Thus, our results may not generalize outside of the United States. We hope that future work can combine web-browsing data from around the globe, e.g., [7, 51, 96, 106, 114, 123, 184, 185], with web scraping to take a user-centric approach to understanding topical and linguistic differences in misinformation and hard news exposure around the world. Furthermore, our results are from a specific period in time: the 2020 U.S. Presidential Election. Thus, the temporal validity [115] of our results may not be generalizable to other elections or periods of lower political salience. Also, our findings only relate to exposure to misinformation and hard news. Therefore, our findings may not translate into understanding the potential effects of misinformation, particularly because past work has shown that exposure to misinformation does not necessarily equate to belief in misinformation [166].

6 Conclusion

The study takes a user-centric approach to understand the linguistic features and topical patterns of misinformation consumed during the 2020 U.S. Presidential Election. Our findings show that misinformation is generally more readable, infused with moral language, and exhibits higher negative sentiment compared to hard news. Furthermore, we found that the linguistic characteristics of misinformation are not uniform but vary significantly across issue contexts. We also observed individual heterogeneity in the consumption of misinformation: older adults tend to consume more misinformation related to COVID-19 and health, with content showing more negative sentiment and fewer moral terms. Republicans interact with misinformation that features greater negative sentiment and more moral language, while paying less attention to health topics and more to social and political issues. These results highlight the importance of tailoring interventions to combat misinformation based on specific user characteristics.

Acknowledgments

The authors thank participants at the 2025 International Communication Association conference, as well as members of the Stanford Social Media Lab, including Cid Decatur, Bingxu Han, Angela Lee, Sunny Liu, Ryan C. Moore, Lindsay Popowski, Ronald Robertson, Aysu Sarigul, Will Schulz, Ilana Shumsky, Debbie Siegel, Anja Stevic, Sarah Wu, and Harry Yaojun Yan, for feedback on this work. We are also grateful to Deepak Kumar for his help with scraping the web pages. We would also like to thank the participants who provided the data to make this study possible. This project was funded by a MURI award no. W911NF-20-1-0252 and support from Stanford's Freeman Spogli Institute for International Studies. R.D. was supported by graduate fellowship awards from Knight-Hennessy Scholars and Stanford Data Science at Stanford University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting hoaxes, frauds, and deception in writing style online. In *2012 IEEE symposium on security and privacy*. IEEE, 461–475.

[2] Zhila Aghajari, Eric PS Baumer, and Dominic DiFranzo. 2023. Reviewing interventions to address misinformation: the need to expand our vision beyond an individualistic focus. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–34.

[3] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. 2020. Fake news detection using machine learning ensemble methods. *Complexity* 2020, 1 (2020), 8885861.

[4] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & politics* 6, 2 (2019), 2053168019848554.

[5] Jennifer Allen, Antonio A Arechar, Gordon Pennycook, and David G Rand. 2021. Scaling up fact-checking using the wisdom of crowds. *Science advances* 7, 36 (2021), eabf4393.

[6] Jennifer Allen, Duncan J Watts, and David G Rand. 2024. Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science* 384, 6699 (2024), eadk3451.

[7] Sacha Altay, Rasmus Kleis Nielsen, and Richard Fletcher. 2022. Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media* 2 (2022).

[8] American National Election Studies. 2025. ANES 2024 Time Series Study Preliminary Release: Pre-Election Data [dataset and documentation]. https://www.electionstudies.org. February 19, 2025 version.

[9] Scott Appling, Amy Bruckman, and Munmun De Choudhury. 2022. Reactions to Fact Checking. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–17.

[10] Paul Atwell, Fernando B Mello, and Simon Chaucharda. [n. d.]. Reducing Falsehoods at the Source: An experiment incentivizing Brazilian political elites to avoid online misinformation. ([n. d.]).

[11] Ibrahim Al Azher, Venkata Devesh Reddy Seethi, Akhil Pandey Akella, and Hamed Alhoori. 2024. LimTopic: LLM-based Topic Modeling and Text Summarization for Analyzing Scientific Articles limitations. In *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*. 1–12.

[12] Bence Bago, David G Rand, and Gordon Pennycook. 2020. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General* 149, 8 (2020), 1608–1613. doi:10.1037/xge0000729

[13] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[14] Jonathan Baron and John T Jost. 2019. False equivalence: Are liberals and conservatives in the United States equally biased? *Perspectives on Psychological Science* 14, 2 (2019), 292–303.

[15] Roy F Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D Vohs. 2001. Bad is stronger than good. *Review of general psychology* 5, 4 (2001), 323–370.

[16] Zhaleh Beheshti, Daryush Nejadansari, and Hossein Barati. 2020. The relationship between emotional intelligence, lexical diversity and the syntactic complexity of EFL learners' written productions. *Journal of Modern Research in English Language Studies* 7, 1 (2020), 133–161.

[17] Yochai Benkler, Robert Faris, and Hal Roberts. 2018. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

[18] Nicolas Berlinski, Margaret Doyle, Andrew M Guess, Gabrielle Levy, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, and Jason Reifler. 2023. The effects of unsubstantiated claims of voter fraud on confidence in elections. *Journal of Experimental Political Science* 10, 1 (2023), 34–49.

[19] Charles F Bond Jr and Bella M DePaulo. 2006. Accuracy of deception judgments. *Personality and social psychology Review* 10, 3 (2006), 214–234.

[20] Porismita Borah, Yan Su, Xizhu Xiao, and Danielle Ka Lai Lee. 2022. Incidental news exposure and COVID-19 misperceptions: A moderated-mediation model. *Computers in Human Behavior* 129 (2022), 107173.

[21] Shelley Boulianne and Adam Shehata. 2022. Age differences in online news consumption and online political expression in the United States, United Kingdom, and France. *The International Journal of Press/Politics* 27, 3 (2022), 763–783.

[22] Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology* 15 (2013), 209–227.

[23] William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* 114, 28 (2017), 7313–7318.

[24] Nadia M Brashier and Daniel L Schacter. 2020. Aging in an era of fake news. *Current directions in psychological science* 29, 3 (2020), 316–323.

[25] Megan A Brown, James Bisbee, Angela Lai, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2022. Echo chambers, rabbit holes, and algorithmic bias: How YouTube recommends content to real users. *Available at SSRN 4114905* (2022).

[26] Roberto Cabeza, Nicole D Anderson, Jill K Locantore, and Anthony R McIntosh. 2002. Aging gracefully: compensatory brain activity in high-performing older adults. *Neuroimage* 17, 3 (2002), 1394–1402.

[27] Matt Carlson, Sue Robinson, and Seth C Lewis. 2021. Digital press criticism: The symbolic dimensions of Donald Trump's assault on US journalists as the "enemy of the people". *Digital Journalism* 9, 6 (2021), 737–754.

[28] Carlos Carrasco-Farré. 2022. The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities and Social Sciences Communications* 9, 1 (2022), 1–18.

[29] Laura L Carstensen. 2006. The influence of a sense of time on human development. *Science* 312, 5782 (2006), 1913–1915.

[30] Andreu Casas and Nora Webb Williams. 2019. Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly* 72, 2 (2019), 360–375.

[31] Marina Charquero-Ballester, Jessica G Walter, Ida A Nissen, and Anja Bechmann. 2021. Different types of COVID-19 misinformation have different emotional valence on Twitter. *Big Data & Society* 8, 2 (2021), 20539517211041279.

[32] Robert Chew, John Bollenbacher, Michael Wenger, Jessica Speer, and Annice Kim. 2023. LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint arXiv:2306.14924* (2023).

[33] Matthew Childs, Cody Buntain, Milo Z. Trujillo, and Benjamin D. Horne. 2022. Characterizing Youtube and Bitchute content and mobilizers during us election fraud discussions on twitter. In *Proceedings of the 14th ACM Web Science Conference 2022*. 250–259.

[34] Yuwei Chuai, Haoye Tian, Nicolas Pröllochs, and Gabriele Lenzini. 2024. Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter? *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–52.

[35] Yuwei Chuai and Jichang Zhao. 2022. Anger can make fake news viral online. *Frontiers in Physics* 10 (2022), 970174.

[36] Mauro Conti, Daniele Lain, Riccardo Lazzeretti, Giulio Lovisotto, and Walter Quattrociocchi. 2017. It's always April fools' day!: On the difficulty of social network misinformation classification via propagation features. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE, 1–6.

[37] Ross Dahlke and Jeffrey Hancock. 2022. Effects of Misinformation on Election Beliefs: Disentangling Motivated Reasoning from Selective Exposure. doi:10.31219/osf.io/325tn

[38] Ross Dahlke, Deepak Kumar, Zakir Durumeric, and Jeffrey T Hancock. 2023. Quantifying the Systematic Bias in the Accessibility and Inaccessibility of Web Scraping Content from URL-Logged Web-Browsing Digital Trace Data. *Social Science Computer Review* (2023), 08944393231218214.

[39] Kristen D Deppe, Frank J Gonzalez, Jayme L Neiman, Carly Jacobs, Jackson Pahlke, Kevin B Smith, and John R Hibbing. 2015. Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment and Decision making* 10, 4 (2015), 314–331.

[40] Peter H Ditto, Brittany S Liu, Cory J Clark, Sean P Wojcik, Eric E Chen, Rebecca H Grady, Jared B Celniker, and Joanne F Zinger. 2019. At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science* 14, 2 (2019), 273–291.

[41] Joan Donovan, Becca Lewis, and Brian Friedberg. 2018. Parallel Ports. *Post-digital cultures of the far right: Online actions and offline consequences in Europe and the US, Transcript Verlag* (2018), 49–65.

[42] Paul Dourish. 2006. Implications for design. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 541–550.

[43] Melissa-Ellen Dowling. 2023. Far-right populism in alt-tech: A challenge for democracy? *new media & society* (2023), 14614448231205889.
[44] Adam Enders, Christina Farhart, Joanne Miller, Joseph Uscinski, Kyle Saunders, and Hugo Drochon. 2023. Are republicans and conservatives more likely to believe conspiracy theories? *Political Behavior* 45, 4 (2023), 2001–2024.
[45] Cengiz Erisen, Sung-youn Kim, Elif Erisen, and Yüksel Alper Ecevit. 2025. Source Credibility and Issue Type in Misinformation Correction: Foundations of Trust-building in Two Non-WEIRD Countries. (2025).
[46] Marina Ernst. 2024. Identifying textual disinformation using large language models. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. 453–456.
[47] Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science* 8, 3 (2013), 223–241.
[48] FactCheck.org. [n. d.]. Our Process. https://www.factcheck.org/our-process/ Accessed: 2024-10-29.
[49] Lisa Fazio. 2020. Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review* 1, 2 (2020).
[50] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.
[51] Richard Fletcher, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2023. More diverse, more politically varied: How social media, search engines and aggregators shape news repertoires in the United Kingdom. *New Media & Society* 25, 8 (2023), 2118–2139.
[52] Jeremy A Frimer, Reihane Boghrati, Jonathan Haidt, Jesse Graham, and Morteza Dehgani. 2019. Moral foundations dictionary for linguistic analyses 2.0. *Unpublished manuscript* (2019).
[53] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political polarization in online news consumption. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 152–162.
[54] R Kelly Garrett and Robert M Bond. 2021. Conservatives' susceptibility to political misperceptions. *Science Advances* 7, 23 (2021), eabf1234.
[55] Michael Geers, Helen Fischer, Stephan Lewandowsky, and Stefan M Herzog. 2024. The political (a) symmetry of metacognitive insight into detecting misinformation. *Journal of Experimental Psychology: General* 153, 8 (2024), 1961.
[56] Amira Ghenai and Yelena Mejova. 2018. Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–20.
[57] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences* 120, 30 (2023), e2305016120.
[58] Meritxell González Bermúdez. 2015. An analysis of twitter corpora and the differences between formal and colloquial tweets. In *Proceedings of the Tweet Translation Workshop 2015*. CEUR-WS. org, 1–7.
[59] Jesse Graham, Jonathan Haidt, and Brian A Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology* 96, 5 (2009), 1029.
[60] Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, suppl_1 (2004), 5228–5235.
[61] Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. 2022. *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.
[62] Andrew M Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (2023), 398–404.
[63] Andrew M Guess and Kevin Munger. 2023. Digital literacy and online political behavior. *Political science research and methods* 11, 1 (2023), 110–128.
[64] Andrew M Guess, Brendan Nyhan, and Jason Reifler. 2020. Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour* 4, 5 (2020), 472–480.
[65] Chen Guo, Nan Zheng, and Chengqi Guo. 2023. Seeing is not believing: a nuanced view of misinformation warning efficacy on video-sharing social media platforms. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–35.
[66] Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. 2021. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems* 117 (2021), 47–58.
[67] Michael Hameleers. 2024. The visual nature of information warfare: the construction of partisan claims on truth and evidence in the context of wars in Ukraine and Israel/Palestine. *Journal of Communication* (2024), jqae045.
[68] Valerie Hauch, Iris Blandón-Gitlin, Jaume Masip, and Siegfried L Sporer. 2015. Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and social psychology Review* 19, 4 (2015), 307–342.
[69] Annett Heft, Eva Mayerhöffer, Susanne Reinhardt, and Curd Knüpfer. 2020. Beyond Breitbart: Comparing right-wing digital news infrastructures in six western democracies. *Policy & internet* 12, 1 (2020), 20–45.

[70] Benjamin Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. 759–766.

[71] Matthew J Hornsey, Matthew Finlayson, Gabrielle Chatwood, and Christopher T Begeny. 2020. Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. *Journal of Experimental Social Psychology* 88 (2020), 103947.

[72] Farnaz Jahanbakhsh, Amy X Zhang, Adam J Berinsky, Gordon Pennycook, David G Rand, and David R Karger. 2021. Exploring lightweight interventions at posting time to reduce the sharing of misinformation on social media. *Proceedings of the ACM on human-computer interaction* 5, CSCW1 (2021), 1–42.

[73] Chenyan Jia, Alexander Boltz, Angie Zhang, Anqing Chen, and Min Kyung Lee. 2022. Understanding effects of algorithmic vs. community label on perceived accuracy of hyper-partisan misinformation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[74] Chenyan Jia, Angela Yuson Lee, Ryan C Moore, Cid Halsey-Steve Decatur, Sunny Xun Liu, and Jeffrey T Hancock. 2024. Collaboration, crowdsourcing, and misinformation. *PNAS nexus* 3, 10 (2024), pgae434.

[75] John T Jost. 2017. Ideological asymmetries and the essence of political psychology. *Political psychology* 38, 2 (2017), 167–208.

[76] Mark Jurkowitz and Amy Mitchell. 2020. Older Americans continue to follow COVID-19 news more closely than younger adults. *Pew Research Center* 22 (2020).

[77] Mona Kasra, Cuihua Shen, and James F O'Brien. 2018. Seeing is believing: How people fail to identify fake images on the web. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6.

[78] John Kerr, Costas Panagopoulos, and Sander Van Der Linden. 2021. Political polarization on COVID-19 pandemic response in the United States. *Personality and individual differences* 179 (2021), 110892.

[79] Bhanu Khetharpal, Anandita Khanooja, Abhishree Verma, Ankita Ankita, and Rishabh Kaushal. 2024. Understanding Influence Operations via Images. In *Companion Publication of the 16th ACM Web Science Conference*. 115–119.

[80] Jana Lasser, Segun T Aroyehun, Fabio Carrella, Almog Simchon, David Garcia, and Stephan Lewandowsky. 2023. From alternative conceptions of honesty to alternative facts in communications by US politicians. *Nature human behaviour* 7, 12 (2023), 2140–2151.

[81] Noëlle Lebernegg, Jakob-Moritz Eberl, Petro Tolochko, and Hajo Boomgaarden. 2024. Do You Speak Disinformation? Computational Detection of Deceptive News-Like Content Using Linguistic and Stylistic Features. *Digital Journalism* (2024), 1–24.

[82] Angela Y Lee, Ryan C Moore, and Jeffrey T Hancock. 2024. Building resilience to misinformation in communities of color: Results from two studies of tailored digital media literacy interventions. *new media & society* (2024), 14614448241227841.

[83] Sian Lee, Aiping Xiong, Haeseung Seo, and Dongwon Lee. 2023. "Fact-checking" fact checkers: A data-driven approach. *Harvard Kennedy School Misinformation Review* (2023).

[84] Or Levi, Pedram Hosseini, Mona Diab, and David A Broniatowski. 2019. Identifying nuances in fake news vs. satire: using semantic and linguistic cues. *arXiv preprint arXiv:1910.01160* (2019).

[85] Timothy R Levine. 2014. Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology* 33, 4 (2014), 378–392.

[86] Ro'ee Levy. 2021. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review* 111, 3 (2021), 831–870.

[87] Stephan Lewandowsky and Sander Van Der Linden. 2021. Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology* 32, 2 (2021), 348–384.

[88] Yingdan Lu and Jennifer Pan. 2021. Capturing clicks: How the Chinese government uses clickbait to compete for visibility. *Political Communication* 38, 1-2 (2021), 23–54.

[89] Jason Lucas, Adaku Uchendu, Michiharu Yamashita, Jooyoung Lee, Shaurya Rohatgi, and Dongwon Lee. 2023. Fighting fire with fire: The dual role of LLMs in crafting and detecting elusive disinformation. *arXiv preprint arXiv:2310.15515* (2023).

[90] Josephine Lukito. 2024. Global Misinformation & Disinformation Special Issue Introduction. *International Journal of Public Opinion Research* 36, 3 (2024), edae030.

[91] Bernhard Lutz, Marc Adam, Stefan Feuerriegel, Nicolas Pröllochs, and Dirk Neumann. 2024. Which linguistic cues make people fall for fake news? A comparison of cognitive and affective processing. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–22.

[92] Benjamin A Lyons. 2022. Why we should rethink the third-person effect: disentangling bias and earned confidence using behavioral data. *Journal of Communication* 72, 5 (2022), 565–577.

[93] Benjamin A Lyons, Jacob M Montgomery, Andrew M Guess, Brendan Nyhan, and Jason Reifler. 2021. Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences* 118,

23 (2021), e2019527118.
[94] Rakoen Maertens, Jon Roozenbeek, Jon S Simons, Stephan Lewandowsky, Vanessa Maturo, Beth Goldberg, Rachel Xu, and Sander van der Linden. 2025. Psychological booster shots targeting memory increase long-term resistance against misinformation. *Nature Communications* 16, 1 (2025), 2062.
[95] Lisa Mekioussa Malki, Dilisha Patel, and Aneesha Singh. 2024. " The Headline Was So Wild That I Had To Check": An Exploration of Women's Encounters With Health Misinformation on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–26.
[96] Frank Mangold, David Schoch, and Sebastian Stier. 2024. Ideological self-selection in online news exposure: Evidence from Europe and the US. *Science Advances* 10, 37 (2024), eadg9287.
[97] Marie-Louise Mares, Anne Bartsch, and James Alex Bonus. 2016. When meaning matters more: Media preferences across the adult life span. *Psychology and Aging* 31, 5 (2016), 513.
[98] David M Markowitz and Jeffrey T Hancock. 2019. Deception and language: The contextual organization of language and deception (COLD) framework. *The Palgrave handbook of deceptive communication* (2019), 193–212.
[99] Cameron Martel, Jennifer Allen, Gordon Pennycook, and David G Rand. 2024. Crowds can effectively identify misinformation at scale. *Perspectives on Psychological Science* 19, 2 (2024), 477–488.
[100] Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications* 5 (2020), 1–20.
[101] Hana Matatov, Mor Naaman, and Ofra Amir. 2022. Stop the [Image] steal: The role and dynamics of visual content in the 2020 US Election Misinformation Campaign. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.
[102] Killian L McLoughlin, William J Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, and MJ Crockett. 2024. Misinformation exploits outrage to spread online. *Science* 386, 6725 (2024), 991–996.
[103] Lindsey Meeks. 2020. Defining the enemy: How Donald Trump frames the news media. *Journalism & Mass Communication Quarterly* 97, 1 (2020), 211–234.
[104] Amy Mitchell, Jeffrey Gottfried, Galen Stocking, Mason Walker, and Sophia Fedeli. 2019. Many Americans say made-up news is a critical problem that needs to be fixed. *Pew Research Center* 5 (2019), 2019.
[105] Rafael A Monteiro, Roney LS Santos, Thiago AS Pardo, Tiago A De Almeida, Evandro ES Ruiz, and Oto A Vale. 2018. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13*. Springer, 324–334.
[106] Camila Mont'Alverne, Amy Ross Arguedas, Sayan Banerjee, Benjamin Toff, Richard Fletcher, and Rasmus Kleis Nielsen. 2024. The electoral misinformation nexus: how news consumption, platform use, and trust in news influence belief in electoral misinformation. *Public Opinion Quarterly* 88, SI (2024), 681–707.
[107] Ryan C Moore, Ross Dahlke, Priyanjana Bengani, and Jeffrey Hancock. 2023. The Consumption of Pink Slime Journalism: Who, What, When, Where, and Why? *OSF preprint* (2023).
[108] Ryan C Moore, Ross Dahlke, Peter Forberg, and Jeffrey Hancock. 2024. The private life of QAnon: A mixed methods investigation of Americans' exposure to QAnon content on the web. (2024).
[109] Ryan C Moore, Ross Dahlke, and Jeffrey T Hancock. 2023. Exposure to untrustworthy websites in the 2020 US election. *Nature Human Behaviour* 7, 7 (2023), 1096–1105.
[110] Ryan C Moore and Jeffrey T Hancock. 2022. A digital media literacy intervention for older adults improves resilience to fake news. *Scientific reports* 12, 1 (2022), 6008.
[111] Mohsen Mosleh, Qi Yang, Tauhid Zaman, Gordon Pennycook, and David G Rand. 2024. Differences in misinformation sharing can lead to politically asymmetric sanctions. *Nature* (2024), 1–8.
[112] Matthew Motta. 2021. Republicans, not democrats, are more likely to endorse anti-vaccine misinformation. *American Politics Research* 49, 5 (2021), 428–438.
[113] Yida Mu, Chun Dong, Kalina Bontcheva, and Xingyi Song. 2024. Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 10160–10171.
[114] Subhayan Mukerjee. 2024. Online news in India: a quantitative appraisal of the digital news consumption landscape in the world's largest democracy (2014–2018). *Information, Communication & Society* (2024), 1–21.
[115] Kevin Munger. 2023. Temporal validity as meta-science. *Research & Politics* 10, 3 (2023), 20531680231187271.
[116] Sean A Munson and Paul Resnick. 2010. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1457–1466.
[117] Usman Naseem, Adam Dunn, Matloob Khushi, and Jinman Kim. [n. d.]. Vaccine Misinformation Detection in X using Cooperative Multimodal Framework. *ACM Multimedia 2024*.
[118] W Russell Neuman, Marion R Just, and Ann N Crigler. 1992. *Common knowledge: News and the construction of political meaning*. University of Chicago Press.

[119] Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin* 29, 5 (2003), 665–675.

[120] Zoe Ziyi Ng, Grace Li, Suzanne Flynn, and W Quin Yow. 2023. How COVID-19 news affect older adults' mental health—Evidence of a positivity bias. *International Journal of Environmental Research and Public Health* 20, 5 (2023), 3950.

[121] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. 677–686.

[122] Finn Årup Nielsen. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903* (2011).

[123] Rasmus Kleis Nielsen and Richard Fletcher. 2022. Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the UK. *Journal of Quantitative Description: Digital Media* 2 (2022).

[124] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[125] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. 2023. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 620, 7972 (2023), 137–144.

[126] Joseph T Ornstein, Elise N Blasingame, and Jake S Truscott. 2023. How to train your stochastic parrot: Large language models for political texts. *Political Science Research and Methods* (2023), 1–18.

[127] Irene V Pasquetto, Eaman Jahani, Shubham Atreja, and Matthew Baum. 2022. Social debunking of misinformation on WhatsApp: the case for strong and in-group ties. *Proceedings of the ACM on human-computer interaction* 6, CSCW1 (2022), 1–35.

[128] Yilang Peng, Yingdan Lu, and Cuihua Shen. 2023. An agenda for studying credibility perceptions of visual misinformation. *Political Communication* 40, 2 (2023), 225–237.

[129] Yilang Peng, Sijia Qian, Yingdan Lu, and Cuihua Shen. 2025. Large Language Model-Informed Feature Discovery Improves Prediction and Interpretation of Credibility Perceptions of Visual Content. *arXiv preprint arXiv:2504.10878* (2025).

[130] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.

[131] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2019. Understanding and reducing the spread of misinformation online. *Unpublished manuscript: https://psyarxiv. com/3n9u8* (2019), 1–84.

[132] Gordon Pennycook, Jonathan A Fugelsang, and Derek J Koehler. 2015. What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive psychology* 80 (2015), 34–72.

[133] Gordon Pennycook and David G Rand. 2019. Cognitive reflection and the 2016 US Presidential election. *Personality and Social Psychology Bulletin* 45, 2 (2019), 224–239.

[134] Gordon Pennycook and David G Rand. 2020. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality* 88, 2 (2020), 185–200.

[135] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104* (2017).

[136] Elizabeth A Phelan, Lynda A Anderson, Andrea Z LaCroix, and Eric B Larson. 2004. Older adults' views of "successful aging"—how do they compare with researchers' definitions? *Journal of the American Geriatrics Society* 52, 2 (2004), 211–216.

[137] Tiziano Piccardi, Martin Saveski, Chenyan Jia, Jeffrey T Hancock, Jeanne L Tsai, and Michael Bernstein. 2024. Social Media Algorithms Can Shape Affective Polarization via Exposure to Antidemocratic Attitudes and Partisan Animosity. *arXiv preprint arXiv:2411.14652* (2024).

[138] Sijia Qian, Cuihua Shen, and Jingwen Zhang. 2023. Fighting cheapfakes: using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *Journal of Computer-Mediated Communication* 28, 1 (2023), zmac024.

[139] Ellen Quintelier. 2007. Differences in political participation between young and old people. *Contemporary politics* 13, 2 (2007), 165–180.

[140] Emilee Rader and Rebecca Gray. 2015. Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 173–182.

[141] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[142] Maria Rae. 2021. Hyperpartisan news: Rethinking the media for populist politics. *New Media & Society* 23, 5 (2021), 1117–1132.
[143] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2931–2937.
[144] Steve Rathje, James K He, Jon Roozenbeek, Jay J Van Bavel, and Sander Van Der Linden. 2022. Social media behavior is associated with vaccine hesitancy. *PNAS nexus* 1, 4 (2022), pgac207.
[145] Steve Rathje, Dan-Mircea Mirea, Ilia Sucholutsky, Raja Marjieh, Claire E Robertson, and Jay J Van Bavel. 2024. GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences* 121, 34 (2024), e2308950121.
[146] Jonathan Rauch. 2021. *The constitution of knowledge: A defense of truth.* Brookings Institution Press.
[147] Andrew E Reed, Larry Chan, and Joseph A Mikels. 2014. Meta-analysis of the age-related positivity effect: age differences in preferences for positive over negative information. *Psychology and aging* 29, 1 (2014), 1.
[148] Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. Structural topic models for open-ended survey responses. *American journal of political science* 58, 4 (2014), 1064–1082.
[149] Ronald E Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing partisan audience bias within google search. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–22.
[150] Jon Roozenbeek and Sander Van der Linden. 2019. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 1 (2019), 1–10.
[151] Jon Roozenbeek, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. Psychological inoculation improves resilience against misinformation on social media. *Science advances* 8, 34 (2022), eabo6254.
[152] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. Encounters with visual misinformation and labels across platforms: An interview and diary study to inform ecosystem approaches to misinformation interventions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.
[153] Erik J Schlicht. 2025. Analyzing the temporal dynamics of linguistic features contained in misinformation. *arXiv preprint arXiv:2503.04786* (2025).
[154] Lanyu Shang, Ziyi Kou, Yang Zhang, and Dong Wang. 2022. A duo-generative approach to explainable multimodal covid-19 misinformation detection. In *Proceedings of the ACM Web Conference 2022*. 3623–3631.
[155] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM transactions on intelligent systems and technology (TIST)* 10, 3 (2019), 1–42.
[156] Elisa Shearer and Amy Mitchell. 2021. News use across social media platforms in 2020. (2021).
[157] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.
[158] Andrea Sipka, Aniko Hannak, and Aleksandra Urman. 2022. Comparing the Language of QAnon-related content on Parler, Gab, and Twitter. In *Proceedings of the 14th ACM Web Science Conference 2022*. 411–421.
[159] Marina Solnyshkina, Radif Zamaletdinov, Ludmila Gorodetskaya, and Azat Gabitov. 2017. Evaluating text complexity and Flesch-Kincaid grade level. *Journal of social studies education research* 8, 3 (2017), 238–248.
[160] Dominik Stammbach, Vilém Zouhar, Alexander Hoyle, Mrinmaya Sachan, and Elliott Ash. [n. d.]. Revisiting Automated Topic Model Evaluation with Large Language Models. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.
[161] Dominik A Stecula and Mark Pickup. 2021. How populism and conservative media fuel conspiracy beliefs about COVID-19 and what it means for COVID-19 behaviors. *Research & Politics* 8, 1 (2021), 2053168021993979.
[162] Natalie Jomini Stroud and Emily Van Duyn. 2023. Curbing the decline of local news by building relationships with the audience. *Journal of Communication* 73, 5 (2023), 452–462.
[163] Mubashir Sultan, Alan N Tump, Nina Ehmann, Philipp Lorenz-Spreen, Ralph Hertwig, Anton Gollwitzer, and Ralf HJM Kurvers. 2024. Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences* 121, 47 (2024), e2409329121.
[164] S Shyam Sundar, Maria D Molina, and Eugene Cho. 2021. Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication* 26, 6 (2021), 301–319.
[165] Henri Tajfel and John C Turner. 1978. Intergroup behavior. *Introducing social psychology* 401, 466 (1978), 149–178.
[166] Christopher K Tokita, Kevin Aslett, William P Godel, Zeve Sanderson, Joshua A Tucker, Jonathan Nagler, Nathaniel Persily, and Richard Bonneau. 2024. Measuring receptivity to misinformation at scale on a social media platform. *PNAS nexus* 3, 10 (2024), pgae396.

[167] Fatemeh Torabi Asr and Maite Taboada. 2019. Big Data and quality data for fake news and misinformation detection. *Big data & society* 6, 1 (2019), 2053951719843310.
[168] Fangjing Tu. 2024. Empowering social media users: nudge toward self-engaged verification for improved truth and sharing discernment. *Journal of Communication* 74, 3 (2024), 225–236.
[169] Fangjing Tu, Zhongdang Pan, and Xinle Jia. 2023. Facts are hard to come by: discerning and sharing factual information on social media. *Journal of Computer-Mediated Communication* 28, 4 (2023), zmad021.
[170] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185, 4157 (1974), 1124–1131.
[171] Joseph Uscinski, Adam Enders, Amanda Diekman, John Funchion, Casey Klofstad, Sandra Kuebler, Manohar Murthi, Kamal Premaratne, Michelle Seelig, Daniel Verdear, et al. 2022. The psychological and political correlates of conspiracy theory beliefs. *Scientific reports* 12, 1 (2022), 21672.
[172] Asahi Ushio. 2023. LM-PPL: Language Model Perplexity. https://github.com/asahi417/lmppl. Accessed: 2025-06-30.
[173] Peter Van Aelst, Fanni Toth, Laia Castro, Václav Štětka, Claes de Vreese, Toril Aalberg, Ana Sofia Cardenal, Nicoleta Corbu, Frank Esser, David Nicolas Hopmann, et al. 2021. Does a crisis change news habits? A comparative study of the effects of COVID-19 on news media use in 17 European countries. *Digital Journalism* 9, 9 (2021), 1208–1238.
[174] Jay J Van Bavel and Andrea Pereira. 2018. The partisan brain: An identity-based model of political belief. *Trends in cognitive sciences* 22, 3 (2018), 213–224.
[175] Sander Van der Linden. 2023. *Foolproof: Why misinformation infects our minds and how to build immunity.* WW Norton & Company.
[176] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science* 359, 6380 (2018), 1146–1151.
[177] Herun Wan, Shangbin Feng, Zhaoxuan Tan, Heng Wang, Yulia Tsvetkov, and Minnan Luo. 2024. Dell: Generating reactions and explanations for llm-based misinformation detection. *arXiv preprint arXiv:2402.10426* (2024).
[178] Xinyu Wang, Jiayi Li, and Sarah Rajtmajer. 2024. Inside the echo chamber: Linguistic underpinnings of misinformation on Twitter. In *Proceedings of the 16th ACM Web Science Conference*. 31–41.
[179] Yuping Wang, Chen Ling, and Gianluca Stringhini. 2023. Understanding the use of images to spread COVID-19 misinformation on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.
[180] Brian E Weeks, Ericka Menchen-Trevino, Christopher Calabrese, Andreu Casas, and Magdalena Wojcieszak. 2023. Partisan media, untrustworthy news sites, and political misperceptions. *New Media & Society* 25, 10 (2023), 2644–2662.
[181] WEF. 2024. *Global Risks Report 2024.* Technical Report. World Economic Forum.
[182] Andrew Westbrook, Daria Kester, and Todd S Braver. 2013. What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PloS one* 8, 7 (2013), e68210.
[183] Florian Wintterlin, Tim Schatto-Eckrodt, Lena Frischlich, Svenja Boberg, Felix Reer, and Thorsten Quandt. 2023. "It's us against them up there": Spreading online disinformation as populist collective action. *Computers in Human Behavior* 146 (2023), 107784.
[184] Magdalena Wojcieszak, Bernhard Clemm von Hohenberg, Andreu Casas, Ericka Menchen-Trevino, Sjifra de Leeuw, Alexandre Gonçalves, and Miriam Boon. 2022. Null effects of news exposure: a test of the (un) desirable effects of a 'news vacation'and 'news binging'. *Humanities and Social Sciences Communications* 9, 1 (2022), 1–10.
[185] Magdalena Wojcieszak, Ericka Menchen-Trevino, Bernhard Clemm von Hohenberg, Sjifra de Leeuw, João Gonçalves, Sam Davidson, and Alexandre Gonçalves. 2024. Non-news websites expose people to more political content than news websites: Evidence from browsing data in three countries. *Political Communication* 41, 1 (2024), 129–151.
[186] Madelyne Xiao and Jonathan Mayer. 2023. The challenges of machine learning for trust and safety: a case study on misinformation detection. *arXiv preprint arXiv:2308.12215* (2023).
[187] Harry Yaojun Yan, Ryan C Moorea, Fangjing Tu, and Jeffery T Hancocka. 2025. Detecting Synthetic, Doubting Authentic: AI Attribution Bias for Political Imagery. *OSF Preprints* (2025).
[188] Kai-Cheng Yang, Pranav Goel, Alexi Quintana-Mathé, Luke Horgan, Stefan D McCabe, Nir Grinberg, Kenneth Joseph, and David Lazer. 2025. DomainDemo: a dataset of domain-sharing activities among different demographic groups on Twitter. *arXiv preprint arXiv:2501.09035* (2025).
[189] Jinyi Ye, Luca Luceri, and Emilio Ferrara. 2024. Auditing Political Exposure Bias: Algorithmic Amplification on Twitter/X Approaching the 2024 US Presidential Election. *arXiv preprint arXiv:2411.01852* (2024).
[190] Himanshu Zade, Megan Woodruff, Erika Johnson, Mariah Stanley, Zhennan Zhou, Minh Tu Huynh, Alissa Elizabeth Acheson, Gary Hsieh, and Kate Starbird. 2023. Tweet Trajectory and AMPS-based Contextual Cues can Help Users Identify Misinformation. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–27.
[191] Alvin Zhou, Tian Yang, and Sandra González-Bailón. 2023. The puzzle of misinformation: Exposure to unreliable content in the United States is higher among the better informed. *new media & society* (2023), 14614448231196863.
[192] Cheng Zhou, Kai Li, and Yanhong Lu. 2021. Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness. *Information Processing & Management* 58, 6 (2021),

102679.

[193] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[194] Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–40.

[195] Yanmengqian Zhou and Lijiang Shen. 2022. Confirmation bias and the persistence of misinformation on climate change. *Communication Research* 49, 4 (2022), 500–523.

A Prompt for Categorizing Topics Using GPT-4o mini

Article: {article}

Classify the above news article into one of the following topics.

Topic: ["Social_Issues", "US_Electoral_Politics", "COVID-19", "Health", "General_News", "Missing"]

Topic definitions:

* Social_Issues: News focusing on current societal debates and topics, such as religion, ideologies, anti-muslim, race, gender and sexual orientation, immigration, extremist and conspiracy movements, QAnon, Black Lives Matter, protests, civil unrest, law enforcement, national gun control, global elites, new world order, social justice, freedom of speech, religious freedom, international relations, and non-covid related economy.

* US_Electoral_Politics: News related to the US election cycle, such as coverage of candidates (e.g., Donald Trump and Joe Biden), campaigns, voting processes, election outcomes, election poll, and presidential debate.

* COVID-19: News related to the COVID-19 pandemic, such as COVID-19 health and safety guidelines, the health and economic impacts of COVID-19, COVID-19 vaccination, government responses to COVID-19, parental medical informed consent rights, COVID-19 lockdowns, anti-lockdown protest, COVID-19 & Medical Censorship, and misinformation surrounding the virus.

* Health: News on non-Covid health topics, such as non-traditional health practices, diets, nutrition, supplements, remedies, health treatments or theories, and healthcare.

* General_News: Other news that does not fit into the previous categories, such as climate, weather, sports, business, tech, education, local crime, isolated incidence of crime or one-off incidence of shooting, and other topics.

* Missing: Text that indicates the page is not reachable.

Output should be a json containing a probability distribution over the 6 topics.

B Case Studies: Topic-Specific Comparisons

These case studies provide examples of the linguistic differences between misinformation and hard news identified in our quantitative analysis. Through paired comparisons within each topic, we demonstrate how our statistical findings about readability, sentiment, moral language, and perplexity manifest in actual misinformation and hard news content.

B.1 Health Misinformation vs. Hard News

Table 5. Health Topic Comparisons

Misinformation Example	Hard News Example
"Did you know that AstraZeneca, manufacturer of two blockbuster breast cancer drugs (one of which is classified as a known human carcinogen), was the originator of Breast Cancer Awareness Month? Why is it, do you think, that during Breast Cancer Awareness Month (BCAM) you never hear the word "carcinogen" mentioned, but are barraged a million times over by the word "cure"?"	*"Everyone's body stores fat differently. The lower belly tends to be a place where fat collects for many people. This is because of several factors: genetics, diet, inflammation, and lifestyle choices. Doctors recommend a balanced approach to weight management."*

These health-related articles demonstrate key linguistic differences identified in our analysis. The misinformation article's accusatory tone ("known human carcinogen", "never hear the word carcinogen") exemplifies the significantly more negative sentiment we found in health misinformation compared to hard news ($\gamma = -3.26$, $p < 0.001$). The contrast in tone between the misinformation's confrontational questions and the hard news article's neutral, educational approach aligns with our broader findings about linguistic differences in health coverage.

B.2 COVID-19 Misinformation vs. Hard News

Table 6. COVID-19 Topic Comparisons

Misinformation Example	Hard News Example
"Coronavirus: Sweden's Senior Epidemiologist: Wearing Face Masks Is Very Dangerous. Having face masks and then think[ing] you can crowd your buses or your shopping malls – that's definitely a mistake."	*"MUSKEGON, MI – Inside the walls of Mercy Hospital in Muskegon, employees who are terrified and untrained are being pulled in to care for a cascade of coronavirus patients who have filled the hospital."*

These COVID-19 articles demonstrate several key patterns from our analysis. The misinformation's negative framing ("very dangerous", "definitely a mistake") reflects our finding that COVID-19 misinformation shows more negative sentiment than COVID-19 hard news ($\gamma = -1.55$, $p < 0.001$). The articles' different writing styles—from the misinformation's simple declarative statements to the hard news article's detailed scene-setting and context—reflect our finding that COVID-19 misinformation shows lower perplexity than COVID-19 hard news ($\gamma = -81.08$, $p < 0.001$).

Table 7. Social Issues Comparisons

Misinformation Example	Hard News Example
"Black Lives Matter: Two Very Different Meanings. The problem with radical groups being taken purely at face value. More than ever this year, we have seen people all over the world throwing their unquestioning support behind radical political groups after taking their names purely at face value."	*"Nineteen years ago Friday, attacks by the Islamist terrorist cult Al Qaeda took place on American soil, followed by conspiracy theories that the CIA bombed the World Trade Towers and the Pentagon. These have been thoroughly debunked, but they have still flourished, as Al Qaeda did itself until the U.S. took the threat seriously."*

B.3 Social Issues Misinformation vs. Hard News

These social issues articles demonstrate key linguistic patterns from our analysis. While both articles discuss controversial topics with similar emotional weight, our findings indicate no significant sentiment difference in social issues content ($\gamma = -0.08$, $p = 0.779$). Both articles' frequent use of moral language ("radical groups", "terrorist cult") illustrates why social issues content shows the highest moral language across all topics ($\beta = 1.92$, $p < 0.001$). The contrast between the repetitive phrasing of misinformation and the varied structure of hard news articles reflects our finding of lower perplexity in social issues misinformation ($\gamma = -81.52$, $p < 0.001$).

B.4 U.S. Electoral Politics Misinformation vs. Hard News

Table 8. Electoral Politics Comparisons

Misinformation Example	Hard News Example
"It's not clear what Trump's goal is, other than to give an FU to the country that dared to toss him out of office. What is for certain is that it's not going to benefit the Pentagon or the U.S. CNN notes that neither Bossie nor Lewandowski have served in the military or have any apparent experience with the defense industry."	*"Democratic presidential nominee Joe Biden's lead over Donald Trump has surged to a record 17 points as the US election enters its final sprint, an Opinium Research and Guardian opinion poll shows."*

These articles both cover aspects of the 2020 U.S. presidential transition and demonstrate key linguistic differences. The misinformation article's informal language ("FU") and emotional accusations ("dared to toss him out") reflect our finding that political misinformation shows more negative sentiment than hard news ($\gamma = -0.83$, $p = 0.002$). The contrast between the colloquial style of misinformation and the structured reporting of hard news, with polling data and institutional citations, exemplifies our finding that political misinformation shows lower perplexity than hard news ($\gamma = -91.08$, $p < 0.001$).

B.5 General News Misinformation vs. Hard News

These examples illustrate differences between misinformation and hard news reporting styles. The misinformation's sensational style ("Hollywood Horror", "Secret Satanic Cult!") and dramatic

Table 9. General News Comparisons

Misinformation Example	Hard News Example
"Hollywood Horror: Oscar Winner's Murder-Suicide Linked to Secret Satanic Cult! Exclusive photos from crime scene show occult symbols painted in blood!"	*"Actor Todd Nance, founding member of rock band Widespread Panic, dies at 57. Manager confirms complications from long-term illness. Memorial services scheduled for Friday."*

language ("painted in blood") demonstrates more complex linguistic patterns compared to the hard news article's straightforward reporting ("Manager confirms", "Memorial services scheduled"), aligning with our finding that general news misinformation shows significantly higher perplexity than hard news ($\beta = 76.60$, $p < 0.001$).

C Examining Self-Reported Bias

One potential limitation of our analysis is self-report bias, in which individuals may inaccurately report their demographic and political partisanship. We took a three-part approach to evaluate the potential of self-report bias. First, we assessed the face validity of the self-reported partisanship and age measure by examining the question's wording. Second, we assessed the measures' convergent validity by examining their correlations with other theoretically correlated self-reported measures. Third, we analyzed the correlation between self-reported partisanship and a behavioral measure of revealed preferences of partisan media consumption. Altogether, our results suggested that our self-reported partisanship measure was generally valid and was correlated with other self-reported and behavioral measures.

First, we examined self-reported political affiliation. We reviewed the wording of the survey question for political affiliation to assess face validity. It was measured using questions from well-established national surveys, including the American National Election Survey and Pew Research, as described in Section 3.3. The professional survey agency YouGov, who distributed our survey, collected these responses repeatedly at 30-day intervals; the repeated measures account for its reliability and accuracy over time.

Second, we assessed the convergent validity of our measure. Convergent validity refers to whether theoretically related measures are actually empirically related. To examine this, we assessed the correlation between political partisanship and both political ideology and voting choices. Theoretically, it is expected that Republicans should be ideologically conservative and more likely to vote for Republican candidates, such as Donald Trump in the 2020 U.S. Presidential Election. Our results supported this: participants who self-identified as Republican reported a mean political ideology score of 1.07 on a 5-point scale ranging from -2 (very liberal) to 2 (very conservative), whereas those who identified as non-Republican had a mean score of -0.73. This difference was statistically significant (t = 26.679, p <.001). We also found that participants who self-reported being a Republican were more likely to self-report voting for Donald Trump in the 2020 U.S. Presidential Election than Democrats ($\chi^2(1, N = 719) = 334.28$, p <.001). These results showed that the self-reported data aligned with theoretical expectations.

While we found that our measure of self-reported partisanship was correlated with another related self-reported measure, we also examined the extent to which self-reported partisanship was correlated with the behavioral measure of partisan media consumption. To do so, we joined our web-browsing data to partisan bias metrics developed by Robertson et al. [149]. These partisan bias

scores cover over 19,000 domains and provide a score between -1 and 1 for each domain. A score of -1 means that the domain was exclusively visited by Democrats. A score of 1 means the domain was visited exclusively by Republicans. Scores around 0 were visited by about equal proportions of Democrats and Republicans. Importantly, the partisanship of these individuals was measured via U.S. voter registration information in a voter file.

In our analysis, we averaged the partisan bias scores of the domains each individual visited to calculate their individual partisan media consumption score. These scores can be seen as reflecting individuals' revealed partisan preferences (at least in media consumption) and is a measure we can validate against the self-reported partisanship measure. We found that the two measures were significantly correlated. Self-reported Republicans had a mean partisan media score of 0.134, meaning that they visited more Republican-leaning domains, and non-Republicans had a mean bias score of 0.022, indicating they were less likely to visit the Republican-leaning domains. The distributions of these scores were significantly different across the two groups ($t = 152.98$, $p < .001$).

The other critical self-reported measure in our study is age. The exact wording of the question was described in the Measures section. Our approach of asking participants for their year of birth is comparatively more robust than asking for age directly, as it reduces imprecise or inaccurate responses. Additionally, the professional survey agency YouGov updates this information every 27 months. For participants who have been in the panel for more than one year, multiple birth year measures are available, allowing for cross-validation of the data. Moreover, our results showed that older participants consume more hard news than younger participants, which aligned with theoretical expectations and supported the measure's convergent validity.

While we primarily assess the face validity and stability of self-reported age and partisanship, we have only the behavioral measure to evaluate the convergent validity of partisanship. Hopefully, future work can provide additional behavioral variables that can be used in checking convergent validity for age, e.g., [188].

Received October 2024; revised April 2025; accepted August 2025