What drives engagement with disinformation on social media? A Social Identity Approach

Anonymous ACL submission

Abstract

Research on user engagement with disinformation has expanded significantly in recent years, yet the mechanisms that make implausible narratives persuasive remain unclear. While emotion, source credibility, and information structure have been found to influence disinformation endorsement, this study argues that linguistic and collective identity plays a significant role in shaping engagement. It posits that sharing and interacting with disinformation—via comments, reactions, and reposts—functions as a discursive practice encoding both collective agency and cultural-linguistic identity. Using a dataset of 3,885 tweets in English (EN), French (FR), and Spanish (ES), the study examines whether language choice and call-to-action expressions (e.g., mobilization phrases) drive engagement. Statistical analysis, topic modelling, and linguistic analyses reveal notable cross-linguistic differences: Spanish-speaking users engage significantly more with directive language, while English and French users exhibit distinct, less predictable interaction patterns. These findings challenge the assumption that mobilization rhetoric universally boosts engagement and suggest that audience responsiveness to disinformation-based engagement is shaped by linguistic and cultural context. These results therefore underscore the need to consider linguistic and cultural variations in disinformation research as well as tailored communication strategies in disinformation mitigation and digital content dissemination.

1 Introduction

Online disinformation—understood here as the deliberate dissemination of false or misleading content with the potential to cause public harm (Tucker et al., 2018)—has surged in recent years, particularly in the wake of the COVID-19 crisis. Widely acknowledged as a major threat to public and individual safety, substantial research has examined its structure (Van Prooijen and Douglas, 2018; Van Prooijen and Van Vugt, 2018), spread (Bonnevie et al., 2021), impact (Simms et al., 2020; Stabile et al., 2019; Chen et al., 2020), and content (Wiggins, 2023; Demata et al., 2022; Fallis, 2009). Scholars have also investigated the role of conspiracy theories and disinformation in shaping public perceptions and decision-making (Chen et al., 2021; Yagi et al., 2024), as well as the influence of network structures and user interactions in amplifying disinformation (Quintana et al., 2022; Gunaratne et al., 2019). More recently, attention has turned to understanding the deeper mechanisms through which disinformation persuades individuals to accept unlikely or false narratives, giving more importance to the cognitive and identity-based factors that may explain why individuals engage with disinformation (Reddi et al., 2023; Bastick, 2021; Butter and Knight, 2020).

Despite the rapid growth of this field, however, the fundamental reasons behind user engagement with disinformation remain poorly understood. Some researchers argue that this gap stems from contradictory and fragmented findings (Birchall and Knight, 2022; Kirchner and Reuter, 2020), while others point to the overly functionalist approach to disinformation, treating it as merely 'the opposite of true' and overlooking its cognitive and subjective dimensions (Viola, 2025; Reddi et al., 2023; Bastick, 2021). This often leads to disinformation consumers being dismissed as irrational or paranoid actors and to counter-measures being mostly inefficient (Alava et al., 2017; Conway, 2017; Johnson, 2018; Mølmen and Ravndal, 2021; Reicher and Haslam, 2016).

This study contends that a critical gap in the literature lies in the failure to address the role of linguistic identity in driving engagement with disinformation. It argues that engagement with disinformation is not merely an act of passive belief but an active discursive process, where individuals construct and negotiate their cultural identities.

It further argues that interacting with disinformation—through shares, comments, reactions, and reposts—constitutes a discursive practice that equally encodes collective agency. This behaviour would be shaped by perceived injustice, disillusionment with mainstream media, and the affordances of alternative information ecosystems (Wintterlin et al., 2023).

These arguments are supported by two key considerations. First, much of the current research on disinformation remains disproportionately focused on a small subset of industrialized democracies, particularly the United States and the United Kingdom (Bajaj, 2024). A study by Seo & Faris (Seo and Faris, 2021) found that 62.8% of empirical studies published in communication journals between 2015 and 2020 relied on U.S.-based data (p. 1166). Scholars such as Bajaj (Bajaj, 2024) highlight that disinformation is not a universal phenomenon and that this geographical bias distorts our understanding of its cultural dimensions. Consequently, mitigation efforts that ignore these cultural dynamics risk being ineffective.

Second, disinformation has been linked to citizens' decrease of trust in mainstream media and other sources of authoritative information (MacFarquhar, 2016; Lewis and Marwick, 2017; Allcott and Gentzkow, 2017), mistrust of establishment political figures and institutions and increased acceptance of – or indeed support for – fringe, anti-establishment or radical actors and movements (Beauchamp, 2019; Amlinger and Nachtwey, 2022; Reichardt, 2022). Research on the 2020 health crisis, for example, has demonstrated that COVID-19 disinformation motivated individuals to protest by offering a sense of agency and empowerment (Reichardt, 2022; Amlinger and Nachtwey, 2022; Birchall and Knight, 2022). Thus, according to this view, even unlikely or improbable disinformation narratives succeed in mobilizing individuals by fostering a belief in their capacity to effect change. In this sense, online participation—expressing dissent, signing petitions, or sharing content—would function as a low-cost form of collective action (Brunsting and Postmes, 2002). Through social media, users can signal group membership, express opposition to elites, and reinforce a shared identity, transforming engagement with disinformation into a performative act of resistance. Online activities such as sharing or liking content offers an easy and effective way for people to express dissent with others or to demonstrate their belonging to a group. Following on these considerations, this study conceptualizes engagement with disinformation as a form of online collective action, driven by group identity and opposition to elite narratives.

Further, this study advocates for a broader, comparative, and multilingual approach in disinformation research that investigates linguistic and cultural identity-based differential experiences of disinformation. To this aim, the paper analyses 3,885 disinformation tweets in English (EN), Spanish (ES), and French (FR). Using statistics, topic modelling, and linguistic analysis, the study examines how linguistic features, topical themes, and call-to-action expressions shape user engagement, using metrics such as likes, retweets, quotes, shares, and replies. The results emphasize the influence of linguistic and cultural identity factors in shaping individuals' susceptibility to persuasive disinformation narratives. They also offer a nuanced perspective that examines the shared symbols and thematic patterns that resonate with social media users in complex environments, such as the Internet.

2 User engagement, disinformation, and collective action

The scholarly literature on user engagement in online and social media contexts has approached this phenomenon through various conceptual frameworks and methodologies. Engagement is often conceptualized as user-initiated actions that contribute to value co-creation, as proposed by Brodie et al. (Brodie et al., 2013). This broad definition underscores the interplay between behavioural, cognitive, and emotional dimensions of engagement, emphasizing the need to explore its motivations and nuances. Shao (Shao, 2009) categorized user interaction into three primary behaviours: consumption (viewing and reading), participation (interacting with content), and production (creating and uploading content). Following this framework, researchers have examined how engagement manifests across platforms, particularly differentiating between active participation and passive consumption. On Facebook and YouTube, active engagement involves actions such as liking, commenting, and sharing, whereas passive engagement consists of clicking, watching, or hovering over content (Kaur et al., 2019; Khan, 2017). On X (formerly Twitter), active engagement further includes retweeting, quoting, and following (Chen, 2011). Studies also highlight the prevalence of pas-

sive users (often called 'lurkers'), who primarily consume content without actively engaging, comprising up to 90% of users in many online communities (Nonnecke and Preece, 1999; Preece et al., 2004). This contrast between active contributors and passive consumers underscores the need to understand what motivates users to actively engage with content, particularly disinformation.

Recent research has therefore explored the drivers of engagement with disinformation and fake news on social media. Emotionally charged content has been identified as one of the strongest amplifiers of engagement, with sensationalized headlines and narratives strategically crafted to trigger emotional responses such as fear, anger, and anticipation, thus encouraging interaction and dissemination (Horner et al., 2023). Additionally, visual elements seem to play a significant role in enhancing credibility and audience response (Cao et al., 2020). Features such as clickbait, emotionally charged language, and references to specific individuals, organizations, or events further heighten emotional resonance, thereby boosting engagement (Ali et al., 2023). Other factors influencing the likelihood of sharing disinformation would include fear of missing out, source credibility, information quality, cognitive overload, and social media fatigue (Kumar et al., 2020; Islam et al., 2020). These content strategies capitalize on psychological triggers to grab attention, manipulate perceptions, and enhance virality.

Furthermore, the role of content creators in the spread of disinformation seems to be critical (Vilella et al., 2022). Audiences would be more likely to trust and engage with content from sources they perceive as authoritative and credible (Dennis et al., 2023). This would explain why disinformation spreaders often exploit trust by mimicking legitimate sources or leveraging the reputations of influential figures (Housholder and LaMarre, 2014). Conversely, established content creators with verified accounts and large followings benefit from increased credibility, making their content more likely to be accepted and widely shared (Dubois et al., 2020). Research has indeed shown that emotionally provocative disinformation—particularly fear-inducing, anger-driven, or surprising content—spreads more rapidly when disseminated by trusted users with significant audiences (Fan and Lederman, 2018).

This study builds upon these findings and integrates the Social Identity Theory (SIT) (Tajfel and Turner, 2004) with the Social Identity Model of Collective Action (SIMCA) (Tajfel and Turner, 2004; Turner, 1991; van Zomeren et al., 2008) to provide a theoretical lens for understanding engagement with disinformation. The SIT framework claims that users engage with content that reinforces their group identity (e.g., linguistic communities, national identity, ideological affiliations). According to this view, engagement is driven by in-group/out-group dynamics, where users support content aligned with their identity and oppose conflicting narratives. Given the absence of specific multilingual and multicultural frameworks for user engagement, SIT provides valuable insights into why users are more likely to engage with content that reflects their linguistic, cultural, or political identity, reinforcing their sense of group belonging in online spaces.

The SIMCA framework posits that collective action is driven by three interrelated psychological mechanisms: social identity, perceived injustice, and anger combined with participatory efficacy. Applied to disinformation, this model helps explain why users engage with misleading narratives not just as passive consumers but as active participants in a form of digital collective action. According to SIMCA, group identification fosters a sense of belonging, reinforcing shared beliefs and in-group solidarity. Hence, individuals who engage with and disseminate disinformation often align themselves with a social group opposed to mainstream narratives (Wintterlin et al., 2023).

Many disinformation narratives frame mainstream institutions (e.g., governments, media, academia) as corrupt or oppressive, fuelling distrust and resentment (Reichardt, 2022). SIMCA suggests that perceived injustice strengthens motivation for collective action, which in the digital sphere translates into engaging with and sharing disinformation as an act of defiance. Finally, emotionally charged content—particularly anger-driven messages—has been shown to increase disinformation engagement(Horner et al., 2023). According to SIMCA, anger acts as a mobilizing force, prompting individuals to take action, while participatory efficacy (the belief that engagement can effect change) encourages users to share disinformation as a means of collective resistance.

By situating disinformation engagement within the SIT and SIMCA frameworks, this study argues that language and culture shape how users interpret and interact with misleading narratives. Emotions

such as fear, anger, frustration, and distrust are deeply influenced by cultural values, lived experiences, and social identities (Viola, 2025). While social media may amply these emotional responses (Bonnevie et al., 2021; Puri et al., 2020), reducing the phenomenon of disinformation to a mere consequence of digital platforms would be an oversimplification (Nguyen and Catalan-Matamoros, 2022). Engagement with disinformation is on the contrary a complex socio-cognitive process, driven by individual traits (e.g., cultural background, personal beliefs, and values) (Buturoiu et al., 2021; Hornsey et al., 2023), intergroup identity dynamics (Cookson et al., 2021; Turner, 1991), and broader sociopolitical conditions such as populism (Demata et al., 2022) and declining trust in institutions (Nguyen and Catalan-Matamoros, 2020).

To investigate these dynamics, this study examines the relationship between language, identity, and topical themes in shaping engagement with disinformation. The findings contribute to a data-driven understanding of how disinformation operates as a tool of collective agency, offering insights into the role of shared discourse in reinforcing identity and mobilizing online action.

3 Data and methodology

The data-set covers a total duration of 394 days spanning a time period from 1 August 2022 to 30 August 2023. It contains 3,885 tweets in 3 languages (English - EN, Spanish -ES, French - FR) and several attributes such as the tweet texts, the hashtags, likes, replies, retweets, shares, and quotes count. Posts were selected for the three chosen languages and extracted from a larger multilingual data-set that had been previously pseudomysed and tagged for disinformation content within the context of a larger study. The final working dataset had 1,933 in Spanish, 1,437 in English, and 516 in French. Figure 1 displays the tweets' distribution per language. This study posits that analyzing user engagement with the selected posts provides quantifiable insights into the dissemination and reception of disinformation. Here, engagement is operationalized as the cumulative sum of likes, shares, retweets, quotes, and replies associated with each tweet. We apply descriptive statistics and Kruskal-Wallis tests to first determine if engagement patterns are language-dependent.

Following the calculation of engagement levels, the analysis employs Latent Dirichlet Allocation

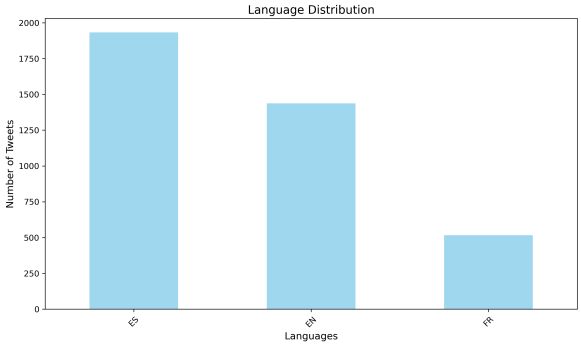

Figure 1: Language distribution in the working dataset

(LDA) topic modelling (TM) (Blei et al., 2003) to examine the thematic structure of the tweets. TM is a computational statistical technique used to identify linguistic patterns in large text datasets. Grounded in distributional semantics theory (Harris, 1954), it operates on the premise that clusters of words convey collective meanings, forming distinct topics. By analyzing the correlation between topic contributions generated through LDA and engagement metrics, this study aims to determine which thematic elements drive audience interaction and dissemination of content in the three languages. Additionally, in order to analyse the effect of collective action in driving engagement with disinformation, we categorized posts based on the presence of call-to-action expressions (e.g., 'Resist the reset', 'donotcomply'), exclamation points, and imperative verb structures (e.g., *'Protégeons nos enfants!'* - let's protect our children, 'noalaagenda2030').

4 Analysis

4.1 Statistical analysis

We first examine possible language-specific differences that can indicate that audience responses vary significantly according to linguistic factors. Conversely, if weak or no correlation is found, this may indicate that other factors (e.g., topic relevance, collective action) might play a larger role in driving engagement. The results are displayed in Figure 2.

The Pearson correlation coefficient is 0.087369 suggesting a weak positive relationship between language and engagement. This would mean that language alone does not strongly influence engagement. However, slight tendencies can be observed across the analysed languages. As shown in Figure 2, engagement is quite scattered across all three languages, with some tweets in EN and FR having high engagement (exceeding 1,500 engagements)

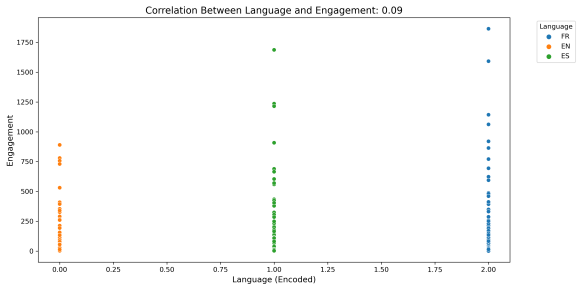

Figure 2: Correlation between language and engagement

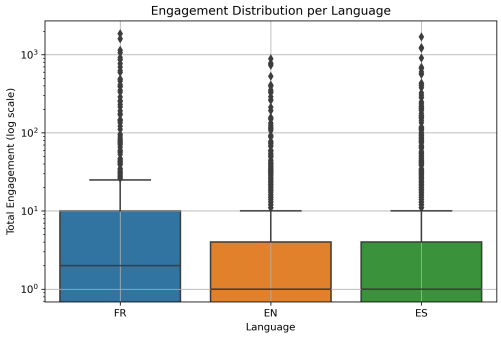

Figure 3: Engagement distribution per language

and indicating possible viral content. Similarly, ES seems to have a number of high-engagement tweets but the engagement is more evenly distributed. As the test has evidenced that the data is not normally distributed, we perform a Kruskal-Wallis test. This will help clarify if engagement patterns are language-dependent. The results are displayed in Figure 3 and Table 1.

Statistic	p-value
41.91	7.91×10^{-10}

Table 1: Kruskal-Wallis Test Results

The statistic value (41.91) indicates a substantial difference in the engagement distributions across languages. Moreover, the extremely low p-value indicates that the engagement levels differ significantly across languages, suggesting that at least one language has a higher or lower engagement level compared to the others. Moreover, although the median engagement is low for all languages, the interquartile range (IQR) is wider for FR, suggesting that French tweets tend to have more variation in engagement. The results also show outliers above the upper whiskers confirming that a few tweets go viral in all the languages. This may indicate that, while engagement does differ between languages, there may also be other factors that explain more variance, which we investigate below.

We now perform pairwise comparisons using the Mann-Whitney U tests to confirm different levels of engagement between languages. Table 2 shows the results.

Comparison	Statistic	p-value
EN vs. FR	302108.0	1.09×10^{-10}
EN vs. ES	1334035.5	0.0415
FR vs. ES	571069.0	1.74×10^{-7}

Table 2: Pairwise Language Engagement Comparison

The very low p-value between EN and FR suggests that English and French tweets have significantly different engagement levels; moreover, the statistic supports a large rank difference, meaning that one language systematically receives more engagement than the other. As observed in previous findings, French tweets receive more engagement than English tweets. The test also shows a moderate difference between EN and ES, but the difference is less pronounced than in the previous pairwise comparison. Finally, there is a significant difference between FR and ES, confirming earlier observations of higher FR engagement and overall confirming that engagement levels significantly differ between languages.

4.2 Topic modelling

This part of the analysis performs topic modelling to capture thematic elements that drive audience interaction and potential patterns and discontinuities across languages. Table 3 displays the different words per topic in each language, whereas Figures 4, 5, and 6 show the most influential words in terms of engagement per language and topic. The results show thematic patterns across the languages: discussions related to health, especially the Covid-19 pandemic and vaccines, are found regardless of the language. The topics also focus on anti-establishment narratives (e.g., ‘elites’), conspiracy theories (e.g., ‘depopulationagenda’, ‘greatreset’, ‘coverup’), climate change, as well as organizations (e.g., ‘WHO’, ‘wef’) and political figures (e.g., ‘Macron’). The term ‘agenda2030’ in particular appears as the top word in all five topics in each language, making it the most central theme across all analysed tweets. The narrative suggests strong opposition to the United Nations Sustainable Development Goals (SDGs), often framed as a globalist

plot to control populations. In EN, related terms are ‘digitalid’, ‘wef’, ‘nwo’, ‘newworldorder’, In FR, ‘odd’ (*Objectifs de Développement Durable*), ‘antivax’. In SP ‘globalismo’, ‘noalaagenda2030’, ‘dictadurasanitaria’.

Notable differences can also be observed. Unlike English and Spanish, French discussions are more domestically framed rather than centred on globalist conspiracies: Macron appears in multiple topics, reflecting high engagement with political distrust in the French government. Moreover, terms like ‘mensonges’ (lies), ‘décès’ (deaths) suggest mistrust and preoccupations over health issues. In Spanish, tweets use stronger mobilization language. Topics include terms like ‘noalaagenda2030’ (no to Agenda 2030), ‘guerra’ (war), and ‘dictadurasanitaria’ (health dictatorship). The framing is therefore more action-oriented, positioning disinformation within a movement for resistance against government and international organizations (ONU, globalistas). Moreover, unlike in EN and FR, Spanish tweets use more emotionally charged words like ‘plandemia’ (pandemic + conspiracy), ‘feliz’ (happy), and ‘verdad’ (truth).

In terms of engagement with specific words, differences can also be observed based on the language. In EN, words related to vaccines and health (‘vaccine’, ‘hpv’), conspiracy theories (e.g., ‘greatreset’), and organizations (e.g., ‘WHO’) generate the highest engagement. This suggests that tweets focusing on institutional distrust, surveillance, global governance, and loss of individual freedoms generate the highest engagement. FR exhibits high engagement linked to health and anti-vaccination discussions (e.g., ‘antivax’) and political figures (e.g., ‘macron’). Such political narratives may reflect a strong distrust of institutions. ES engagement seems to be predominantly driven by anti-establishment narratives (e.g. ‘dictadurasanitaria’), emotions (e.g., ‘feliz’), and COVID-related conspiracies (e.g., ‘plandemia’). These results may signal strong resistance to COVID-19 measures and government-imposed restrictions. Such mistrust echoes themes in English and French but with more anti-globalist rhetoric.

4.3 Collective action

This part of the analysis now investigates the role of collective action language in user engagement with disinformation tweets. The methodology follows a linguistic and statistical approach to identify and analyse tweets that contain collective action expressions as identified in the previous stages of the analysis and additional imperative language and exclamatory sentences. The research hypothesis is that perceived injustice strengthens motivation for collective action, which in an online environment would translate into engaging with and sharing disinformation as an act of defiance and regained agency. We define collective action language as phrases that mobilise users towards participation or resistance. This includes explicit call-to-action phrases (e.g., ‘do not comply’, ‘resist the reset’, ‘mobilízate), imperative verbs (e.g., ‘join’, ‘wake up’, ’protégez’, ‘resiste’), and exclamation marks, which intensify urgency in messages. The data-set was tagged to distinguish tweets with and without collective action expressions and language-specific trends were analysed. The results are displayed in Figure 7.

In English, tweets without collective action expressions receive higher engagement (11.44) than those with it (7.40). This result suggests that call-to-action phrases do not significantly enhance engagement in English-speaking contexts. Users may respond more to informational or ideological content rather than mobilization attempts. In Spanish, on the contrary, tweets with collective action expressions receive significantly higher engagement (24.06) than those without (13.43). Although the

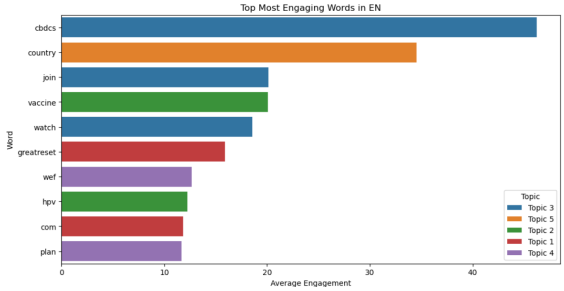

Figure 4: Top 10 high-engagement words in English across topics

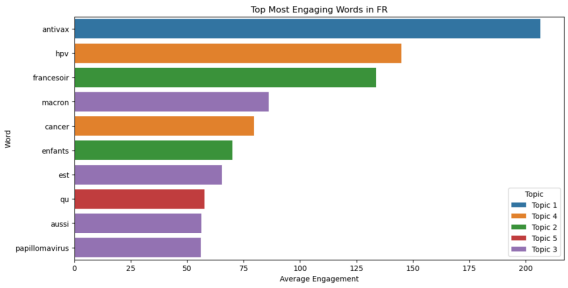

Figure 5: Top 10 high-engagement words in French across topics

Language	Topic	Top 10 Words
EN	Topic 1	agenda2030, com, wef, climatescam, nwo, greatreset, gardasil, 15minutecities, newworldorder, digitalid
	Topic 2	agenda2030, gardasil, climatescam, hpv, vaccine, time, ll, wef, happy, new
	Topic 3	agenda2030, video, wef, gardasil, piped, hpv, cbdcs, watch, donotcomply, join
	Topic 4	agenda2030, agenda, wef, wefpuppets, world, need, plan, canada, 2030, government
	Topic 5	gardasil, agenda2030, thegreatreset, wef, don, coming, org, country, class, wake
FR	Topic 1	gardasil, agenda2030, 2023, est, vaccin, com, odd, vaccination, antivax, rentrée
	Topic 2	gardasil, effets, secondaires, graves, agenda2030, vaccins, francesoir, enfants, décès, macron
	Topic 3	gardasil, papillomavirus, com, est, macron, vaccin, agenda2030, aussi, vaccination, mais
	Topic 4	gardasil, effetssecondaires, enfants, vaccin, papillomavirus, témoignage, contre, cancer, vacciner, hpv
	Topic 5	gardasil, vaccin, papillomavirus, macron, vaccination, fr, macrondestitution, agenda2030, merck, qu
ES	Topic 1	agenda2030, gente, solo, quiere, covid, verdad, pro, ya, nos, guerra
	Topic 2	agenda2030, ods, desarrollo, agenda, sostenible, 2030, objetivos, qué, sociedad, onu
	Topic 3	agenda2030, video, piped, si, tiene, agenda, género, bien, años, esa
	Topic 4	agenda2030, te, globalismo, globalistas, nom, chile, parte, cambio, nos, dictadurasanitaria
	Topic 5	agenda2030, gardasil, ly, plandemia, buff, noalaagenda2030, serás, global, nada, feliz

Table 3: Topic Modeling Results with High-Engagement Words

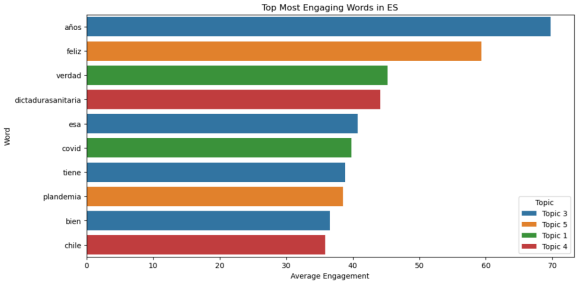

Figure 6: Top 10 high-engagement words in Spanish across topics

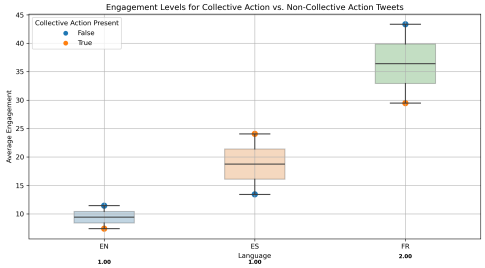

Figure 7: Top 10 high-engagement words in Spanish across topics

maximum engagement for collective action tweets (1,236) is slightly lower than for non-collective action tweets (1,687), the average is much higher. This indicates that while some Spanish tweets without collective action perform well, those with it consistently attract higher engagement. The result suggests that Spanish-speaking users engage more with mobilization messages than English-speaking users, pointing to a stronger collective action culture. Finally, French tweets receive the highest engagement across all languages (43.33 average engagement) regardless of collective action phrases. High engagement is likely driven by political and legal discussions rather than action-based messaging.

5 Discussion

The statistic, topic modelling, and linguistic analysis revealed cross-linguistic variations of engagement with disinformation, showing that language moderately influences engagement with disinfor-

mation. However, since certain topics consistently attracted higher engagement, regardless of language or call-to-action expressions, topics appeared to be the strongest predictors of engagement. This suggests content relevance plays a larger role than linguistic factors alone. Some languages may have a higher baseline engagement (i.e., French), but within each language, topics matter more.

Collective action expressions seem to have the least impact overall. Even though in Spanish tweets showed higher engagement when containing mobilization language, collective action language (imperatives, exclamations, call-to-action phrases) did not significantly increase engagement on the whole. This suggests that mobilization language may matter in some cultures, but the effect is not universal. Thus, to drive engagement, topics are far more important.

Cross-language content engagement patterns could be observed. Vaccine skepticism and health-related disinformation were highly engaging in all three languages, particularly focusing on vaccine safety, mandates, and pharmaceutical distrust. At the same time, political corruption and conspiracy narratives were consistently engaging across all languages but varied in framing. In English, the focus was more against institutions (e.g., WEF, ONU), whereas in French was more on vaccine mandates, side effects, and corruption, and in Spanish the frames were more against globalist elites.

Finally, in terms of psychological drivers of engagement, topics that evoked fear of government overreach, health risks, or loss of freedoms had the highest engagement. However, despite anti-globalist resistance narratives being found across the three languages, tweets with collective action expressions received lower engagement, suggesting perceived injustice does not necessarily strengthen motivation for collective action, at least not in English or French.

6 Conclusion

This study examined the relationship between language, topic, and collective action expressions in driving engagement with disinformation across English (EN), French (FR), and Spanish (ES) tweets. By leveraging statistical, topic modelling, and linguistic analysis, the findings provide empirical evidence that engagement with disinformation is not uniform but shaped by linguistic and cultural identity factors. The results indicate that topic is the strongest predictor of engagement, with certain themes—such as vaccine skepticism, political corruption, and anti-globalist narratives—consistently attracting high interaction. Language plays a moderate role, as French tweets received the highest overall engagement, but within each language, topics dictated engagement levels more than language alone.

Contrary to expectations, collective action expressions (e.g., mobilization phrases, imperatives, exclamation points) had limited influence on engagement. While Spanish tweets with call-to-action expressions saw increased engagement, English and French tweets did not exhibit similar trends. This suggests that mobilization language resonates more strongly within certain cultural contexts but is not a universal driver of engagement.

The study contributes to a growing body of literature emphasizing the need to consider linguistic and cultural variations in disinformation research. It reinforces the complex interplay of language, topic, and identity in shaping online engagement with disinformation. While mobilization language alone does not consistently drive engagement, the findings highlight the role of ideological alignment, emotional resonance, and in-group identity in reinforcing disinformation narratives.

7 Limitations

While this study provides valuable insights, several limitations must be acknowledged. The dataset consists of tweets from X, which may not be representative of broader social media engagement patterns on Facebook, YouTube, or TikTok. Future studies could investigate how different platforms' algorithmic amplification patterns alter engagement dynamics. The study primarily analysed explicit collective action expressions (e.g., imperatives, exclamations) but did not account for subtler mobilization strategies, such as narrative framing or emotional appeals (e.g., victimhood narratives, conspiracy rhetoric). Further research should incorporate semantic and qualitative analysis to assess implicit mobilization strategies. Finally, the study examined aggregate engagement trends rather than individual user behaviours. Future research could explore how user characteristics (e.g., bot accounts, political affiliation, network position) influence engagement with disinformation.

Declaration on Generative AI

During the preparation of this work, the author used GPT-4o in order to: Grammar and spelling check. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

Séraphin Alava, Divina Frau-Meigs, Ghayda Hassan, Hasna Hussein, and Yuanyuan Wei. 2017. *Youth and violent extremism on social media: violent extremism on social media: mapping the research*. United Nations Educational, Scientific, and Cultural Organization, Paris. OCLC: 1089113298.

Maged Ali, Lucas Moreira Gomes, Nahed Azab, João Gabriel de Moraes Souza, M. Karim Sorour, and Herbert Kimura. 2023. Panic buying and fake news in urban vs. rural England: A case study of twitter during COVID-19. *Technological Forecasting and Social Change*, 193:122598.

Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2):211–236.

Carolin Amlinger and Oliver Nachtwey. 2022. *Gekränkte Freiheit: Aspekte des libertären Autoritarismus*, originalausgabe, erste auflage edition. Suhrkamp, Berlin.

Shelly Ghai Bajaj. 2024. Digital Disinformation Threats and Ethnocultural Diasporas. In Gitanjali Adlakha-Hutcheon and Candyce Kelshall, editors, *(In)Security: Identifying the Invisible Disruptors of Security*, pages 53–65. Springer Nature Switzerland, Cham.

Zach Bastick. 2021. Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, 116:106633.

Zack Beauchamp. 2019. Social media is rotting democracy from within. *Vox*.

Clare Birchall and Peter Knight. 2022. *Conspiracy Theories in the Time of Covid-19*, 1 edition. Routledge, London.

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

Erika Bonnevie, Allison Gallegos-Jeffrey, Jaclyn Goldbarg, Brian Byrd, and Joseph Smyser. 2021. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *Journal of communication in healthcare*, 14(1):12–19. ISBN: 1753-8068 Publisher: Taylor & Francis.

Roderick J. Brodie, Ana Ilic, Biljana Juric, and Linda Hollebeek. 2013. Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1):105–114.

Suzanne Brunsting and Tom Postmes. 2002. Social Movement Participation in the Digital Age: Predicting Offline and Online Collective Action. *Small Group Research*, 33(5):525–554. Publisher: SAGE Publications Inc.

Michael Butter and Peter Knight, editors. 2020. *Routledge handbook of conspiracy theories*. Routledge, Abingdon, Oxon ; New York, NY.

Raluca Buturoiu, Georgiana Udrea, Denisa-Adriana Oprea, and Nicoleta Corbu. 2021. Who Believes in Conspiracy Theories about the COVID-19 Pandemic in Romania? An Analysis of Conspiracy Theories Believers' Profiles. *Societies*, 11(4):138. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Juan Cao, Peng Qi, Qiang Sheng, Tianyun Yang, Junbo Guo, and Jintao Li. 2020. Exploring the Role of Visual Content in Fake News Detection. In Kai Shu, Suhang Wang, Dongwon Lee, and Huan Liu, editors, *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, pages 141–161. Springer International Publishing, Cham.

Gina Masullo Chen. 2011. Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, 27(2):755–762.

Li Chen, Qi Ling, Tingjia Cao, and Ke Han. 2020. Mislabeled, fragmented, and conspiracy-driven: a content analysis of the social media discourse about the HPV vaccine in China. *Asian Journal of Communication*, 30(6):450–469.

Li Chen, Yafei Zhang, Rachel Young, Xianwei Wu, and Ge Zhu. 2021. Effects of Vaccine-Related Conspiracy Theories on Chinese Young Adults' Perceptions of the HPV Vaccine: An Experimental Study. *Health Communication*, 36(11):1343–1353.

Maura Conway. 2017. Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies in Conflict & Terrorism*, 40(1):77–98. Publisher: Routledge _eprint: https://doi.org/10.1080/1057610X.2016.1157408.

Darel Cookson, Daniel Jolley, Robert C. Dempsey, and Rachel Povey. 2021. "If they believe, then so shall I": Perceived beliefs of the in-group predict conspiracy theory belief. *Group Processes & Intergroup Relations*, 24(5):759–782.

Massimiliano Demata, Virginia Zorzi, and Angela Zottola, editors. 2022. *Conspiracy theory discourses*. Number volume 98 in Discourse approaches to politics, society and culture. John Benjamins Publishing Company, Amsterdam ; Philadelphia.

Alan R. Dennis, Patricia L. Moravec, and Antino Kim. 2023. Search & Verify: Misinformation and source evaluations in Internet search results. *Decision Support Systems*, 171:113976.

Elizabeth Dubois, Sara Minaeian, Ariane Paquet-Labelle, and Simon Beaudry. 2020. Who to Trust on Social Media: How Opinion Leaders and Seekers Avoid Disinformation and Echo Chambers. *Social Media + Society*, 6(2):2056305120913993.

Don Fallis. 2009. A Conceptual Analysis of Disinformation.

Hanmei Fan and Reeva Lederman. 2018. Online health communities: how do community members build the trust required to adopt information and form close relationships? *European Journal of Information Systems*, 27(1):62–89. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0960085X.2017.1390187.

Keith Gunaratne, Eric A. Coomes, and Hourmazd Haghbayan. 2019. Temporal trends in anti-vaccine discourse on Twitter. *Vaccine*, 37(35):4867–4871.

Zellig Sabbatai Harris. 1954. Distributional structure. *Word. Journal of the linguistic circle of New York*, 10(2-3):146–162.

Christy Galletta Horner, Dennis Galletta, Jennifer Crawford, and Abhijeet Shirsat. 2023. Emotions: The Unexplored Fuel of Fake News on Social Media. In *Fake News on the Internet*. Routledge. Num Pages: 28.

Matthew J. Hornsey, Kinga Bierwiaczonek, Kai Sassenberg, and Karen M. Douglas. 2023. Individual, intergroup and nation-level influences on belief in conspiracy theories. *Nature Reviews Psychology*, 2(2):85–97.

Elizabeth E. Housholder and Heather L. LaMarre. 2014. Facebook Politics: Toward a Process Model for Achieving Political Source Credibility Through Social Media. *Journal of Information Technology & Politics*, 11(4):368–382. Publisher: Routledge _eprint: https://doi.org/10.1080/19331681.2014.951753.

A. K. M. Najmul Islam, Samuli Laato, Shamim Talukder, and Erkki Sutinen. 2020. Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. *Technological Forecasting and Social Change*, 159:120201.

Jessica Johnson. 2018. The Self-Radicalization of White Men: "Fake News" and the Affective Networking of Paranoia. *Communication, Culture and Critique*, 11(1):100–115.

Wandeep Kaur, Vimala Balakrishnan, Omer Rana, and Ajantha Sinniah. 2019. Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity. *Telematics and Informatics*, 39:25–36.

M. Laeeq Khan. 2017. Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66:236–247.

Jan Kirchner and Christian Reuter. 2020. Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–27.

Gaurav Kumar, Rishabh Joshi, Jaspreet Singh, and Promod Yenigalla. 2020. AMUSED: A Multi-Stream Vector Representation Method for Use in Natural Dialogue. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 750–758, Marseille, France. European Language Resources Association.

Becca Lewis and Alice E. Marwick. 2017. Media Manipulation and Disinformation Online. Technical report, Data & Society. Publisher: Data & Society Research Institute.

Neil MacFarquhar. 2016. A Powerful Russian Weapon: The Spread of False Stories - The New York Times.

Guri Nordtorp Mølmen and Jacob Aasland Ravndal. 2021. Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. *Behavioral Sciences of Terrorism and Political Aggression*, 0(0):1–25. Publisher: Routledge _eprint: https://doi.org/10.1080/19434472.2021.1993302.

An Nguyen and Daniel Catalan-Matamoros. 2020. Digital Mis/Disinformation and Public Engagment with Health and Science Controversies: Fresh Perspectives from Covid-19. *Media and Communication*, 8(2):323–328.

An Nguyen and Daniel Catalan-Matamoros. 2022. Anti-Vaccine Discourse on Social Media: An Exploratory Audit of Negative Tweets about Vaccines and Their Posters. *Vaccines*, 10(12):2067.

Blair Nonnecke and Jennifer Preece. 1999. Shedding light on lurkers in online communities. *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities, Edinburgh*, 123128.

Jenny Preece, Blair Nonnecke, and Dorine Andrews. 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior*, 20(2):201–223. Publisher: Elsevier.

Neha Puri, Eric A. Coomes, Hourmazd Haghbayan, and Keith Gunaratne. 2020. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics*, 16(11):2586–2593.

Ignacio Ojea Quintana, Ritsaart Reimann, Marc Cheong, Mark Alfano, and Colin Klein. 2022. Polarization and trust in the evolution of vaccine discourse on Twitter during COVID-19. *PLOS ONE*, 17(12):e0277292. Publisher: Public Library of Science.

Madhavi Reddi, Rachel Kuo, and Daniel Kreiss. 2023. Identity propaganda: Racial narratives and disinformation. *New Media & Society*, 25(8):2201–2218. Publisher: SAGE Publications.

Sven Reichardt, editor. 2022. *Die Misstrauensgemeinschaft der "Querdenker": die Corona-Proteste aus kultur- und sozialwissenschaftlicher Perspektive*, sonderausgabe für die bundeszentrale für politische bildung edition. Number Band 10857 in Schriftenreihe. Bundeszentrale für Politische Bildung, Bonn.

Stephen D. Reicher and Alexander S. Haslam. 2016. Fueling Terror: How Extremists Are Made. *Scientific American*.

Hyunjin Seo and Robert Faris. 2021. Comparative Approaches to Mis/Disinformation| Introduction. *International Journal of Communication*, 15(0):8. Number: 0.

Guosong Shao. 2009. Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research*, 19(1):7–25. Publisher: Emerald Group Publishing Limited.

Kate T. Simms, Sharon J. B. Hanley, Megan A. Smith, Adam Keane, and Karen Canfell. 2020. Impact of HPV vaccine hesitancy on cervical cancer in Japan: a modelling study. *The Lancet. Public Health*, 5(4):e223–e234.

Bonnie Stabile, Aubrey Grant, Hemant Purohit, and Kelsey Harris. 2019. Sex, Lies, and Stereotypes: Gendered Implications of Fake News for Women in Politics. *Public Integrity*, 21(5):491–502. Publisher: Routledge _eprint: https://doi.org/10.1080/10999922.2019.1626695.

Henri Tajfel and John C. Turner. 2004. *The Social Identity Theory of Intergroup Behavior*. Political psychology: Key readings. Psychology Press, New York, NY, US. Pages: 293.

Joshua A. Tucker, Andrew Guess, Pablo Barbera, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.

John C. Turner. 1991. *Social influence*. Social influence. Thomson Brooks/Cole Publishing Co, Belmont, CA, US. Pages: xvi, 206.

Jan-Willem Van Prooijen and Mark Van Vugt. 2018. Conspiracy Theories: Evolved Functions and Psychological Mechanisms. *Perspectives on Psychological Science*, 13(6):770–788.

Jan-Willem Van Prooijen and Karen M. Douglas. 2018. Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, 48(7):897–908.

Martijn van Zomeren, Tom Postmes, and Russell Spears. 2008. Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, 134(4):504–535. Place: US Publisher: American Psychological Association.

Salvatore Vilella, Alfonso Semeraro, Daniela Paolotti, and Giancarlo Ruffo. 2022. Measuring user engagement with low credibility media sources in a controversial online debate. *EPJ Data Science*, 11(1):1–23. Number: 1 Publisher: SpringerOpen.

Lorella Viola. 2025. 'Barren lesbians plotting sterilization': gender stereotypes and prejudices in health disinformation narratives, a cross-cultural analysis of social media of the HPV vaccine. In Catherine Tebaldi, Alistair Plum, and Christoph Purschke, editors, *Conspiracy as Genre: Narrative, Power and Circulation*. Bloomsbury Academic, London.

Bradley Wiggins. 2023. 'Nothing Can Stop What's Coming': An analysis of the conspiracy theory discourse on 4chan's /Pol board. *Discourse & Society*, 34(3):381–398. Publisher: SAGE Publications Ltd.

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*, 146:107784.

Asami Yagi, Yutaka Ueda, and Tadashi Kimura. 2024. HPV Vaccine Issues in Japan: A review of our attempts to promote the HPV vaccine and to provide effective evaluation of the problem through social-medical and behavioral-economic perspectives. *Vaccine*, 42(22):125859.

A Example Appendix

This is an appendix.