information

Article

El Niño Discourse and the Limits of Single-Platform Inference

Dmitry Erokhin * and Nadejda Komendantova

International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria; komendan@iiasa.ac.at
* Correspondence: erokhin@iiasa.ac.at

Abstract

Social media studies often rely on one platform while drawing conclusions about online publics more generally. This study tests that inferential move through an event-centered comparison of El Niño discourse across X/Twitter, YouTube, Facebook, Reddit, TikTok, and LinkedIn. The observation window ran from 9 May through 17 May 2026, several days before and after the May 14 El Niño Watch issued by the National Oceanic and Atmospheric Administration (NOAA), which reported an 82 percent probability of El Niño emerging during May to July 2026 and a 96 percent probability of continuation through the 2026 to 2027 Northern Hemisphere winter. The corpus contains 8145 items classified as highly or moderately related to El Niño after platform-specific collection and common annotation. X/Twitter supplies 7075 items, YouTube 864, Facebook 66, Reddit 59, TikTok 50, and LinkedIn 31. Texts were annotated with a shared structured schema covering relevance, sentiment, emotion, topic, stance, likely misinformation, personal experience, humor, calls to action, language, engagement, and length. The results show that platform choice changes the empirical object. X/Twitter appears multilingual, fast-moving, and weather-heavy. YouTube is more negative, humorous, and personally experiential. Facebook is long-form and media/news oriented, with the highest model-flagged likely misinformation rate. Reddit is concentrated around weather concern. TikTok is short, playful, and personal. LinkedIn is small, professional, and mostly informational. These differences caution against generalizing from one platform to social media as a whole unless a study explicitly defines its scope, accounts for platform and genre differences, and recognizes that visible discourse may include organizational, algorithmically amplified, automated, or otherwise inauthentic activity alongside genuine human expression.

Keywords: social media; platform comparison; external validity; El Niño; climate communication; LLM annotation; misinformation

Academic Editor: Nirmalya Thakur

Received: 8 June 2026
Revised: 22 June 2026
Accepted: 23 June 2026
Published: 24 June 2026

1. Introduction

The growth of social media research has made public communication visible at a scale that was difficult to imagine only a generation ago. Digital traces can show when attention gathers around a forecast, how people frame uncertainty, which emotions travel through online conversation, and which claims become visible under a given platform's ranking and interaction logic. Yet scale does not solve the problem of inference. A large dataset from one platform is still a dataset from one platform. The concern has been central to computational social science for more than a decade. Tufekci [1] warned that social media big data often suffers from unclear sampling frames, platform over-reliance, and field effects. Ruths and Pfeffer [2] similarly argued that population and behavioral biases

differ across platforms and are rarely corrected when researchers use social media to study large-scale behavior. Olteanu et al. [3] later organized these concerns into a broader account of social-data bias, covering biases introduced at the source, during processing, and through ethical or access constraints. The practical implication is not that single-platform studies lack value. It is that their claims should remain tied to the communication environment from which the evidence was drawn. This study builds on that concern by asking what kind of public conversation researchers are seeing and what they might be missing when they study El Niño on one platform. More specifically, the study asks how platform choice shapes findings about El Niño discourse across volume, timing, language, topic, sentiment, emotion, stance, misinformation risk, personal experience, humor, calls to action, engagement, and text length.

El Niño is a useful case because it is not native to one social platform or one communicative genre. It can appear as global climate science, local weather forecasting, food security, agricultural risk, disaster preparedness, energy planning, political argument, humor, personal testimony, or media news. The case is also time sensitive. NOAA's Climate Prediction Center issued an El Niño Watch on 14 May 2026, creating a clear public signal inside the observation window [4]. World Meteorological Organization (WMO)'s May 2026 update likewise described El Niño as increasingly likely through the boreal summer and autumn, while noting potential effects on global temperature and rainfall patterns [5]. The Food and Agriculture Organization (FAO) describes El Niño-induced hazards as risks to food security because disrupted rainfall and temperature patterns can affect agriculture and rural livelihoods [6]. These characteristics make El Niño a strong empirical test. If platforms differ even around a common, scientifically defined issue, then researchers should be cautious when translating one-platform findings into claims about social media discourse as a whole.

Cross-platform discussions may differ empirically, not only because platforms contain different users, but because they organize different forms of visibility, expression, and interaction. These differences shape what researchers measure when they measure sentiment, misinformation, topic, or public concern. To answer the research question, this study compares El Niño-related discourse across X/Twitter, YouTube, Facebook, Reddit, TikTok, and LinkedIn during the same event-centered window. By keeping the issue and time period constant, the analysis shows how the platform itself changes the evidence available to researchers. The goal is not to identify one platform as the best source of public discourse, but to show why claims based on a single platform should be treated with care when researchers want to speak about social media more broadly.

2. Literature and Analytical Framework

Social media platforms are not neutral channels for expression. They organize visibility through their affordances such as ranking mechanisms, interaction norms, available media formats, and the imagined audiences they cultivate. Matassi and Boczkowski [7] argue that comparative research across platforms yields stronger descriptive and explanatory power than treating a single setting as the default. Hase et al. [8] show how news organizations tailor stories differently for Facebook, Instagram, TikTok, and Twitter, while Oz et al. [9] demonstrate that perceived platform affordances shape users' willingness to express opinions. From this perspective, external validity becomes a design choice that starts with the selection of platforms. Bucher and Helmond [10] deepen this view by framing affordances as relational, i.e., emerging among technical design, imagined audiences, user practices, commercial ranking systems, and the task at hand. An El Niño forecast can therefore surface as a breaking-news post, a comment beneath a video, a professional update, a meme-like warning, a subreddit thread, or a link to an official advisory, depending on the platform in question.

https://doi.org/10.3390/info17070622

This multiplicity makes the representativeness problem in social media data especially layered. Platform populations differ, people perform differently across settings, and the very infrastructures including APIs, search tools, and algorithmic rankings expose systematically different slices of the same issue. Tufekci [1], Ruths and Pfeffer [2], and Olteanu et al. [3] all caution against conflating platform-specific traces with social reality. Climate communication brings these biases into sharp focus because a single event routinely moves across registers of scientific explanation, disaster preparedness, market speculation, local memory, political accusation, humor, and outright dismissal of expertise. Pearce et al. [11] found a substantial skew toward Twitter in reviews of climate-related social media research and called for wider attention to platforms, publics, and future imaginaries. For El Niño, a platform suited to tracking fast-moving forecast headlines may not be the one where personal weather stories, food-price anxieties, or skeptical reactions are most visible. Moreover, misleading climate content is not confined to plainly false claims. Herasimenka et al. [12] stress that it also includes manipulative framing and the unintentional amplification of deceptive narratives. In the case of El Niño, that could mean premature certainty about event strength, exaggerated causal claims, conflation of El Niño with global warming, or conspiratorial narratives that replace meteorological explanations with artificial-weather stories [13]. The aim of this study is therefore to show what becomes visible when the same issue is observed across several platformed publics.

To make such comparisons tractable, the analysis attends to six linked dimensions. Attention is captured through volume and timing around the May 14 forecast signal. Genre distinguishes posts, comments, professional updates, subreddit threads, and video replies as non-interchangeable communicative units. Issue framing is recorded via topic categories and main topic. Affect is measured through sentiment, dominant emotion, and emotion intensity. Epistemic stance is labeled as informational, concerned, skeptical, humorous, dismissive, or unclear, together with indicators of likely misinformation and misinformation risk. Finally, interactional style draws on personal experience, humor or sarcasm, calls to action, language, engagement, and text length. Differences along these dimensions are evidence that each platform surfaces a distinct version of public conversation.

Comparing such heterogeneous material demands a measurement strategy that is consistent without being naive. This study uses LLM-assisted annotation as a structured aid for descriptive comparison. As Törnberg [14] notes, this requires attention to model choice, prompt design, validation, reproducibility, and ethics. Ziems et al. [15] advocate for evaluating LLMs against social science constructs rather than treating them as generic text engines, and Fang et al. [16] warn that unexamined annotation error can bias downstream inference. The categories reported here should therefore be read as indicators that reveal broad platform patterns. They complement rather than replace close reading, human validation, and theory-led interpretation.

3. Data and Methods

The data window began at 2026-05-09 00:00:00 UTC and ended before 2026-05-18 00:00:00 UTC. Collection was platform-specific because the six platforms do not provide equivalent public access, search interfaces, official API functionality, or comparable units of analysis. X/Twitter, Facebook, Reddit, TikTok, and LinkedIn were collected through paid Apify actors (see Table 1). YouTube was collected through the YouTube Data API v3. The platform search keyword was the string "El Niño". No language filter was applied in the collection design. All languages returned by the collection route were retained for later relevance filtering and annotation. The only study-imposed collection limit was the 9–17 May 2026 UTC observation window.

The window and platform choices were selected for an event-centered rather than population-representative design. The May 9–17 period brackets a clearly identifiable forecast signal, the May 14 NOAA El Niño Watch, while keeping the observation period short enough to reduce the influence of unrelated news cycles, longer seasonal drift, and platform-level changes over time. The six platforms were chosen because they represent different public-facing communication environments such as fast-moving short posts, video-comment spaces, public media and news posts, community discussion threads, short-form video reactions, and professional updates. The aim is to compare how the same issue becomes visible through several platformed collection routes during the same forecast event.

Table 1. Platform-specific collection design.

Platform	Collector or Access Route	Actor/API Used	Search Basis and Unit
X/Twitter	Apify X/Twitter scraper	kaitoeasyapi/twitter-x-data-tweet-scraper-pay-per-result-cheapest	Latest public posts/tweets matching the keyword
YouTube	YouTube Data API v3	YouTube Data API v3 search and comment endpoints	Videos returned for the keyword, followed by comments and replies under those videos
Facebook	Apify Facebook search actor	scrapeforge/facebook-search-posts	Most recent public posts matching the keyword
Reddit	Apify Reddit post-search actor	easyapi/reddit-posts-search-scraper	Public Reddit posts matching the keyword
TikTok	Apify TikTok comment actor	clockworks/tiktok-comments-scraper	Comments under public TikTok videos identified as relevant to the keyword
LinkedIn	Apify LinkedIn actor	harvestapi/linkedin-post-search	Public posts matching the keyword, with comments where available

The choice of collection route was methodological rather than an attempt to privilege one platform over another. YouTube data were collected through the official YouTube Data API because this route provided a documented and workable method for keyword-based video search and comment retrieval within the event window. For X/Twitter, Facebook, Reddit, TikTok, and LinkedIn, paid Apify actors were used to obtain a synchronized cross-platform snapshot, since official platform APIs differ substantially in access requirements, approval procedures, quotas, endpoint availability, historical-search functionality, and available units of analysis.

The Apify-based collections should therefore be understood as third-party public-web collection routes, not as official platform API access, platform authorization, or platform endorsement. The Apify API token authenticated the workflow only with Apify. It was not used as a login credential for X/Twitter, Facebook, Reddit, TikTok, or LinkedIn. The workflow was configured and interpreted as compliant under the following conditions: collection was limited to material that was publicly visible to the collector at the time of access; no private-account sessions, private communications, closed groups, or access-controlled content were intentionally collected; no paywalls, logins, technical access controls, rate limits, or platform security measures were intentionally bypassed; the collection did not use artificial engagement, account automation, impersonation, or interaction with users; and the resulting data were processed only for research purposes, stored securely, and not redistributed as raw platform records.

Given that platform terms, API rules, and automated-collection policies differ across services and may change over time, this study does not claim that Apify-based collection is equivalent to platform-authorized API collection or that third-party collection is

automatically permitted by each platform. Rather, the authors report the collection route transparently and treat the data as public, collector-visible platform traces subject to strict minimization and reporting safeguards. In particular, the analysis is limited to aggregate patterns. Handles, profile identifiers, and direct links are omitted. Searchable verbatim quotations are avoided, rare combinations of attributes are not reported where they could support identification, and raw social media records are treated as sensitive research material rather than publishable examples or redistributable data.

These precautions are necessary because public accessibility does not remove research responsibilities. Social media content may still contain personal data, contextual identifiers, or traceable expressions, especially when usernames, quotations, timestamps, or platform-specific metadata are reproduced. This manuscript therefore distinguishes between public availability and unrestricted reuse, and it presents the Apify route as a pragmatic, bounded, and transparently disclosed collection method whose acceptability depends on the above criteria being met. This approach is consistent with ethics guidance emphasizing caution around identifiability in social media research [17–19] and with public-data scraping guidance that treats collection as most defensible when it concerns publicly available material, avoids confidential or protected content, respects applicable laws and third-party rights, and does not circumvent access controls [20].

After collection, texts were annotated through the OpenAI API using GPT-5.4-mini model. The annotation schema included sentiment, continuous sentiment score from negative to positive, dominant emotion, emotion intensity, topic, topic category, stance toward El Niño, relevance to El Niño, likely misinformation presence and risk, misinformation claim and reason, personal experience, humor or sarcasm, call to action, detected language, and a short summary.

The LLM workflow was designed as a structured annotation aid [21]. Each item was sent with a source-row identifier and returned as one JSON object under a predefined schema. The schema required the same fields across platforms and restricted several variables to enumerated categories, including sentiment, emotion, topic category, stance, relevance, and misinformation risk. The scripts used strict JSON-schema output, matched outputs back to source rows, checked required fields and allowed values where implemented, flagged missing or invalid outputs, retried failed batches, and saved checkpoints. These safeguards reduce formatting errors, improve reproducibility, and limit format drift, but they do not prove the substantive correctness of every label.

Given that social media texts can contain adversarial, malformed, or instruction-like content, the annotation workflow treated each post or comment only as data to be classified, not as an instruction to the model. The prompts directed the model to analyze only the provided text, avoid inventing outside context, use low-risk or unclear labels when claims were ambiguous, and not treat skepticism or political disagreement as misinformation by default. This conservative instruction was intended to avoid inflating the misinformation category by treating uncertainty, humor, political disagreement, or legitimate skepticism as falsehood. The trade-off is that the instrument may under-detect ambiguous skepticism, irony, coded claims, or humor that indirectly conveys misinformation.

For this reason, the misinformation variable is reported as "likely misinformation" or "model-flagged misinformation risk" and not as verified false content. The same caution applies to stance and affect labels, especially for short, sarcastic, multilingual, or context-dependent comments. The structured-output and post-processing safeguards also reduce the risk of instruction leakage from user-generated text, but they are not equivalent to a full jailbreak or prompt-injection audit [22]. This study therefore treats adversarial robustness as a methodological limitation and recommends that future work add dedicated jailbreak-detection, repeated-call stability, and adversarial test-set checks before using LLM labels for stronger causal or prevalence claims.

The item-level examples added in the results are used only as interpretive illustrations of the model-coded categories. They were selected from the platform-level analysis, then shortened, de-identified, and paraphrased where necessary. Handles, profile names, URLs, and other identifying details are not reproduced. This keeps the examples close enough to explain the labels while avoiding the unnecessary exposure of individual users.

The comparative corpus retains items marked as high or medium relevance to El Niño and excludes unrelated or failed analyses where such fields were available. The final retained dataset contains 8145 items, including 7075 from X/Twitter, 864 from YouTube, 66 from Facebook, 59 from Reddit, 50 from TikTok, and 31 from LinkedIn. The single-keyword strategy was a deliberate precision-oriented design choice. This study used the query to keep the issue definition stable across platforms and to compare explicit discussion of the named climate phenomenon during the same forecast event. This choice improves comparability but reduces recall. For example, it may miss posts using local expressions and indirect references to droughts, floods, crops, food prices, preparedness, or weather risk that do not use the exact collection keyword.

The uneven sample sizes were not a decision to privilege X/Twitter and were not intended to represent platform population sizes. They resulted from applying the same short event window, the same explicit keyword anchor, and the available public collection route on each platform. X/Twitter produced far more retained public items under this design, while Facebook, Reddit, TikTok, and LinkedIn yielded smaller visible corpora. These differences likely reflect platform activity, public-search visibility, ranking and recency logic, genre differences, data-access restrictions, and actor/API behavior. This study therefore treats X/Twitter as a higher-volume visible discourse environment in this corpus and treats smaller platform samples as descriptive profiles rather than stable population estimates.

It cannot be assumed that official APIs would have produced approximately 7000 comparable items from every platform. Official APIs differ in access requirements, quotas, rate limits, approval procedures, historical coverage, comment availability, and search functionality. A future equal-N design could extend the time window, broaden the query list, deliberately oversample smaller platforms, or stratify by platform and genre. Such a design would estimate patterns in a constructed balanced sample rather than compare what became visible through synchronized event-window collection routes.

This study does not verify whether each item was produced by a human user, an organizational account, a coordinated campaign, a bot, or an otherwise inauthentic profile [23]. Automated or fake accounts can amplify climate-related frames and may do so differently across platforms [24]. This study therefore interprets the corpus as visible platform discourse. The main cross-platform claim is that the collected platform environments surfaced different patterns of El Niño-related content, affect, stance, and misinformation risk. Some of those differences may reflect human publics, while others may reflect platform ranking, organizational posting, automated amplification, or fake-profile activity.

4. Results

The first finding is the scale of platform inequality (see Figure 1). X/Twitter accounts for 86.9 percent of all retained items and YouTube accounts for 10.6 percent, while Facebook, Reddit, TikTok, and LinkedIn together account for only 2.5 percent. This does not mean that X/Twitter users cared more about El Niño than users elsewhere. It means that, under these accessible public collection routes and search designs, X/Twitter produced far more visible related material during the window. A combined corpus that ignored platform would therefore mostly reproduce X/Twitter dynamics. A single-platform study of YouTube, LinkedIn, or TikTok would produce a very different empirical account.

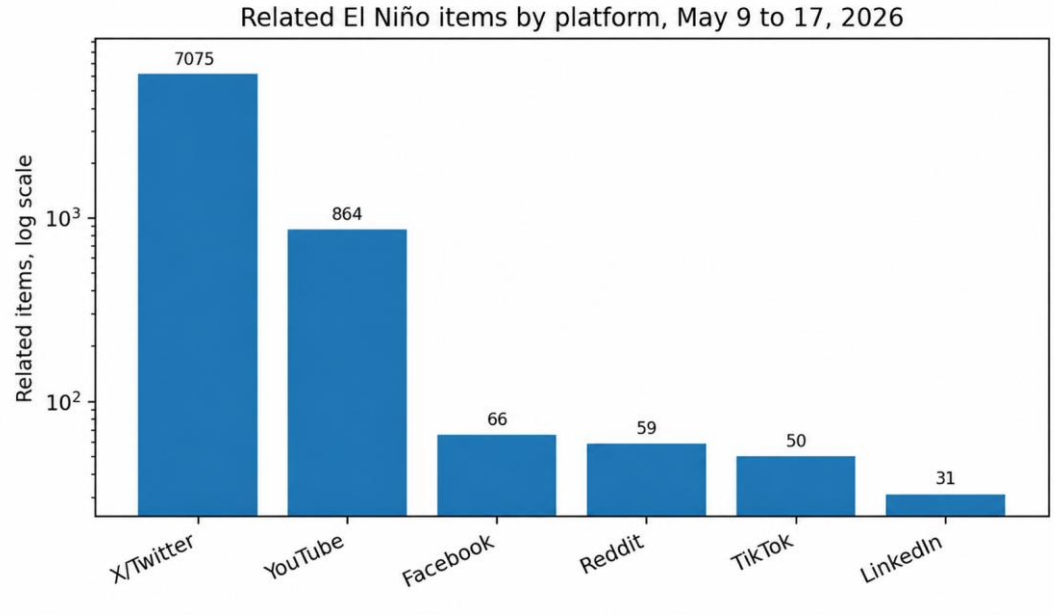

Figure 1. Related El Niño items by platform (9–17 May 2026).

The daily pattern in Figure 2 supports the event-centered interpretation. X/Twitter activity increased from 684 retained items on May 13 to 1161 on May 14 and reached 1614 on May 15. Reddit also shows a visible May 14 increase, with 20 retained items compared with single-digit daily counts on most other days. Facebook, TikTok, and YouTube are more delayed. Facebook rises most strongly on May 17, TikTok appears only from May 15 onward, and YouTube has its largest counts on May 16 and May 17. The same forecast signal therefore did not circulate through each platform at the same speed. Immediate posting, later commenting, video circulation, and public-media sharing created different temporal signatures around one climate event.

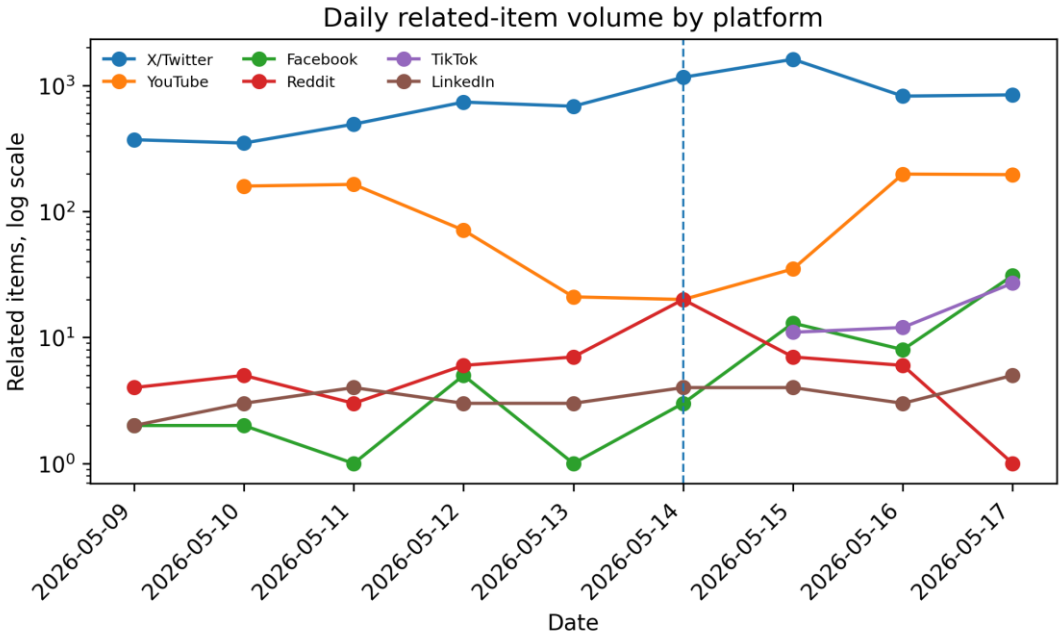

Figure 2. Daily related-item volume by platform. The dashed line corresponds to the May 14 El Niño Watch issued by the NOAA.

The topic and affect results show that the platforms are not merely different in volume (see Table 2). They make different versions of El Niño visible. Weather is the dominant category on X/Twitter, YouTube, Reddit, and TikTok, but its share ranges from 25.9 percent on YouTube to 69.5 percent on Reddit. Facebook differs because media/news is dominant at 53.0 percent. LinkedIn differs again because science/education dominates at 64.5 percent. A researcher who studied Reddit alone would likely describe the issue as weather-centered concern. A researcher who studied LinkedIn alone would describe a professional science and education discourse. A researcher who studied Facebook alone would foreground public media posts and news framing. Each conclusion could be locally valid while remaining incomplete as a general claim about social media.

Table 2. Dominant content profile by platform.

Platform	Dominant Topic	Dominant Emotion	Dominant Stance	Dominant Language Profile
X/Twitter	Weather (37.2%)	Concern (40.3%)	Informational (54.0%)	English 45.4%, Spanish 40.7%, Portuguese 6.0%
YouTube	Weather (25.9%)	Concern (29.7%)	Informational (35.8%)	English (99.4%)
Facebook	Media/news (53.0%)	Concern (47.0%)	Informational (54.5%)	English 63.6%, Spanish 16.7%, Portuguese 6.1%
Reddit	Weather (69.5%)	Concern (49.2%)	Concerned (50.8%)	English (91.5%)
TikTok	Weather (64.0%)	Joy (24.0%)	Informational (32.0%)	English (100.0%)
LinkedIn	Science/education (64.5%)	Trust (25.8%)	Informational (77.4%)	English (90.3%)

Emotion and stance vary just as strongly. Concern is the leading emotion on X/Twitter, YouTube, Facebook, and Reddit, yet the emotional mix differs. YouTube has a notably high anger share, while Reddit concentrates around concern and anticipation. TikTok is led by joy, suggesting a more playful or expressive reaction environment. LinkedIn is led by trust, consistent with its more professional and informational style. Stance follows the same pattern. LinkedIn is overwhelmingly informational, Reddit is majority concerned, X/Twitter combines informational and concerned posts, and YouTube and TikTok are more fragmented across informational, concerned, skeptical, humorous, and dismissive stances. This is the external-validity problem in empirical form where the meaning of El Niño discourse changes with the platform.

Sentiment gives the platform comparison a clearer affective shape. As shown in Table 3, neutral and negative items dominate most of the corpus, but the balance differs across platforms. X/Twitter is primarily neutral, with 62.3 percent neutral items and 34.0 percent negative items, suggesting that much of the discussion circulated as reporting, commentary, or forecast sharing rather than direct emotional reaction. YouTube and Reddit are more negative, with negative items representing 51.6 percent and 49.2 percent of their respective samples, which indicates that comment-centered discussion more often translated the forecast into concern, worry, complaint, or risk language. Facebook is also strongly risk-oriented, with 43.9 percent negative items and concern as the dominant emotion. TikTok is more affectively mixed. Negative and positive items each account for 28.0 percent, while neutral items remain the largest group at 44.0 percent. LinkedIn is the least negative platform and has the highest mean sentiment score, with 19.4 percent positive and 6.5 percent mixed items.

Table 3. Sentiment by platform.

Platform	Mean Sentiment Score	Negative	Neutral	Positive	Mixed
X/Twitter	−0.12	34.0%	62.3%	3.2%	0.5%
YouTube	−0.22	51.6%	35.6%	10.4%	2.3%
Facebook	−0.16	43.9%	56.1%	0.0%	0.0%
Reddit	−0.30	49.2%	47.5%	0.0%	3.4%
TikTok	0.00	28.0%	44.0%	28.0%	0.0%
LinkedIn	0.11	12.9%	61.3%	19.4%	6.5%

Text length captures a different part of the platform effect because it measures the amount of space used to frame the forecast. As shown in Table 4, Facebook posts are by far the longest, with a median of 877 characters and 69.7 percent above 500 characters. LinkedIn is also relatively expansive, with a median of 275 characters. X/Twitter sits between long-form posts and comment environments, with a median of 186 characters and a mean of 257 characters. YouTube comments are shorter, with a median of 107 characters. TikTok comments are shortest, with a median of 49 characters. These differences are analytically important, since longer items can carry more explanation, source framing, qualification, and narrative context. Shorter items more often compress uncertainty into a reaction, joke, warning, question, or fragment of personal memory.

Table 4. Text length by platform.

Platform	Median Chars	IQR Chars	Mean Chars	Median Words	>500 Chars
X/Twitter	186	117–278	257	28	7.4%
YouTube	107	56–203	175	20	5.6%
Facebook	876	446–1414	993	134	69.7%
Reddit	69	38–88	78	12	0.0%
TikTok	49	35–74	65	10	0.0%
LinkedIn	275	146–426	346	37	19.4%

Table 5 compares the available engagement fields within each platform. Facebook has the highest mean engagement at 868.9 per retained item, followed by Reddit at 603.5. TikTok has a high mean of 185.9 because a few comments attracted many likes, although the median is only 3.5. X/Twitter has the largest aggregate engagement total at 442,759, but the median is 1.0, and 39.8 percent of items have zero visible engagement. YouTube contributes many comments, but comment-level engagement is sparse, with a median of zero and a zero-engagement share of 66.4 percent. The engagement results therefore point to platform-specific visibility structures and long-tail attention rather than a universal intensity scale.

Table 5. Engagement by platform-specific available metrics.

Platform	Available Metric	Total	Mean	Median	Max	Zero Engagement
X/Twitter	likes + reposts + replies + quotes	442,759	62.6	1.0	85,937	39.8%
YouTube	comment likes	3974	4.6	0.0	864	66.4%
Facebook	reactions + comments + shares	57,348	868.9	256.0	14,269	4.5%
Reddit	score + comments	35,609	603.5	50.0	6134	1.7%
TikTok	likes + replies	9295	185.9	3.5	5581	24.0%
LinkedIn	likes + comments + shares	313	10.1	1.0	101	35.5%

Language further complicates generalization. X/Twitter is multilingual in the retained corpus, with English at 45.4 percent, Spanish at 40.7 percent, and Portuguese at 6.0 percent. Facebook is mostly English but includes meaningful Spanish and Portuguese shares. YouTube, Reddit, TikTok, and LinkedIn are overwhelmingly English in the retained corpus. This pattern does not prove that non-English discussion was absent from those platforms. It shows what the implemented collection routes and platform environments made visible. An English-dominant YouTube or Reddit study would miss much of the Spanish-language El Niño conversation that appears in X/Twitter. Conversely, an X/Twitter study would make the issue look more multilingual than several other accessible corpora in this project.

The binary and interactional indicators reinforce the same conclusion (see Figure 3). Facebook has the highest likely misinformation rate in the retained corpus at 21.2 percent, followed by LinkedIn at 12.9 percent, X/Twitter at 8.8 percent, TikTok at 8.0 percent, YouTube at 7.1 percent, and Reddit at 3.4 percent. This indicator is model-based and should not be read as external factchecking. Still, it shows that a researcher's estimate of misinformation risk would vary depending on the selected platform. Personal experience is even more platformed. TikTok has the highest personal-experience rate at 24.0 percent, followed by YouTube at 18.9 percent and X/Twitter at 4.1 percent. Humor and sarcasm are highest on YouTube at 19.9 percent and TikTok at 14.0 percent. Calls to action are highest on Facebook at 10.6 percent and X/Twitter at 8.1 percent. Video-comment environments turn El Niño into lived and expressive talk more often, while posting and news environments more often carry informational framing or calls to action.

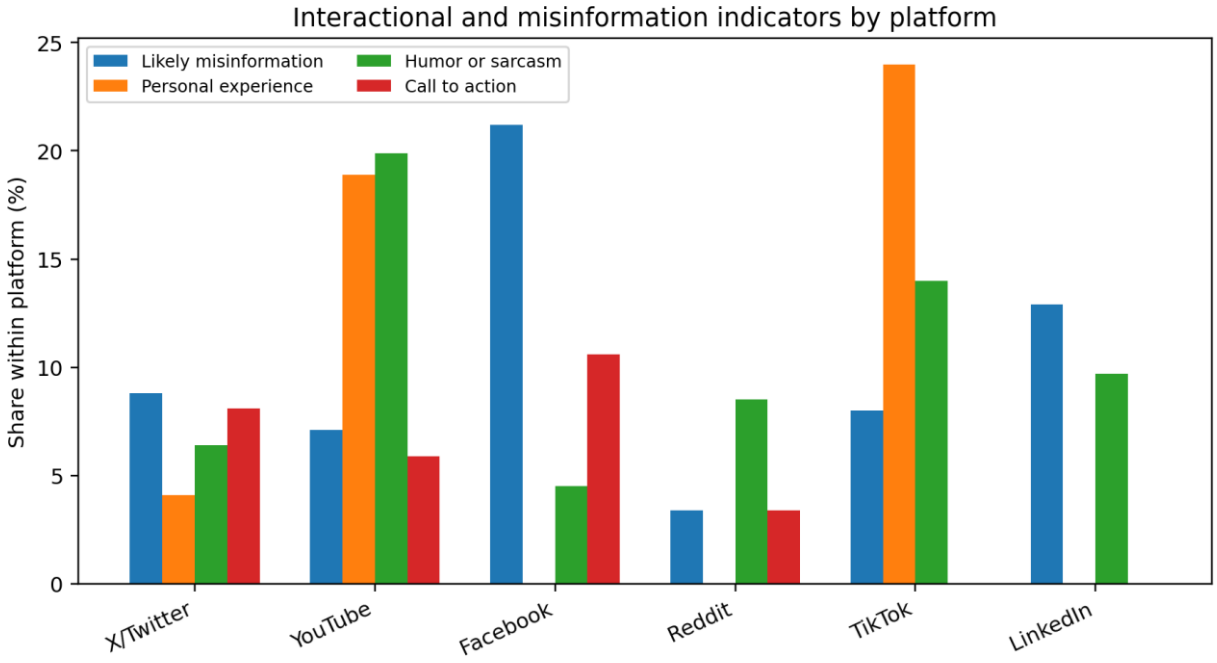

Figure 3. Likely misinformation, personal experience, humor or sarcasm, and call-to-action indicators by platform.

The omnibus tests confirm that the platform differences are not limited to one descriptive indicator (see Table 6). All measured dimensions differ by platform. Among continuous measures, text length and total engagement show the largest effects, while sentiment score differs significantly with a smaller effect. Among categorical measures, personal experience has the largest effect, followed by detected language, dominant emotion, humor or sarcasm, stance, topic category, and sentiment category. Likely misinformation

and calls to action are statistically significant but smaller in effect size. These results should be interpreted as evidence of corpus differentiation, not as estimates of population opinion. The large X/Twitter sample makes small differences detectable, so the substantive reading depends on the direction, magnitude, and interpretive meaning of the platform profiles.

Table 6. Omnibus tests of cross-platform differences.

Domain	Test	Statistic	df	*p*	Effect Size
Sentiment score	Kruskal–Wallis	128.88	5	<0.001	epsilon² = 0.015
Emotion intensity	Kruskal–Wallis	33.10	5	<0.001	epsilon² = 0.003
Text length	Kruskal–Wallis	587.77	5	<0.001	epsilon² = 0.072
Total engagement	Kruskal–Wallis	468.57	5	<0.001	epsilon² = 0.057
Sentiment category	Chi-square	417.57	20	<0.001	Cramer V = 0.113
Dominant emotion	Chi-square	1085.98	55	<0.001	Cramer V = 0.163
Topic category	Chi-square	534.65	60	<0.001	Cramer V = 0.115
Stance toward El Niño	Chi-square	951.63	25	<0.001	Cramer V = 0.153
Detected language	Chi-square	1406.23	290	<0.001	Cramer V = 0.186
Likely misinformation present	Chi-square	18.86	5	0.002	Cramer V = 0.048
Personal experience	Chi-square	345.14	5	<0.001	Cramer V = 0.206
Humor or sarcasm	Chi-square	198.95	5	<0.001	Cramer V = 0.156
Call to action	Chi-square	14.34	5	0.014	Cramer V = 0.042

The aggregate differences become easier to interpret when they are read together with item-level examples. Table 7 shows the kinds of posts, comments, and short-form messages that were placed in each category after de-identification and generalization. The examples clarify that negative often means concern, risk framing, or frustration rather than interpersonal hostility. Likely misinformation can involve unsupported causality, false equivalence, or a shift from probability to certainty, whereas personal memory, humor, requests for facts, preparedness advice, and source-sharing are analytically different forms of forecast talk.

Table 7. Generalized examples of coded categories.

Category	Generalized Examples from the Corpus	Analytical Relevance
Likely misinformation or overclaiming	Items in this category included claims that El Niño was caused by artificial cloud-making, aircraft, solar cycles, underwater volcanic activity, or other unsupported mechanisms, comments treating El Niño as simply another name for global warming, claims that El Niño happens every year in a specific country, and posts that converted a probabilistic forecast into a certain disaster scenario.	This category was broader than obvious denial. It captured unsupported causality, false equivalence, recurring-event misunderstandings, and premature certainty. The important issue was not only whether an item was hostile to climate science, but whether it moved beyond the evidence available in the forecast signal.
Negative affect and risk framing	Items used language such as monster, enormous, frightening, dangerous, catastrophic, disaster, or horror, or expressed frustration that weather communication had become too dramatic. Some items were serious warnings, while others used exaggerated or sarcastic wording around technical forecast discussions.	Negative sentiment therefore captured fear, concern, frustration, sadness, and risk emphasis, not just hostility. It marked the tension between legitimate preparedness language and sensational framing.
Personal experience	Items recalled earlier El Niño years, remembered wet winters and flooding, compared the	These items made a global and probabilistic climate signal local, embodied, and remembered.

	current forecast with past events, reported local snow, rain, heat, drought, or agricultural concern, or expressed preferences for warmer, cooler, wetter, or less drought-prone weather.	Personal experience was analytically separate from information sharing because it translated El Niño into everyday places, seasons, and expectations.
Humor and sarcasm	Items made jokes about billionaires, taxes, data centers, economic crisis language, same-day weather whiplash, or dramatic media style. Some comments used playful exaggeration, eye-rolls, or ironic understatement rather than straightforward agreement or disagreement.	Humor complicated sentiment because it could carry criticism, anxiety, dismissal, or political meaning at the same time. Treating it as noise would hide an important way people process forecast uncertainty and media drama.
Calls to action	Items urged people to prepare, asked communicators to provide direct facts rather than drama, advised leaving high-risk areas, demanded clearer explanation, or proposed adaptation and policy responses such as reflective roofing, emergency planning, water management, or crop-related precautions.	Calls to action showed how forecast talk moved from interpretation to practical response. The category included preparedness, information demands, and policy advocacy rather than a single form of activism.
Science and information sharing	Items summarized official forecasts, explained El Niño basics, shared institutional links, distinguished probability from certainty, described regional effects, or used educational language to clarify what El Niño is and what it does not prove on its own.	The corpus also contained careful informational discourse. This matters because sensationalism and misinformation coexisted with legitimate explanation and source-sharing within the same broad conversation.

These examples show why category labels should be interpreted relationally. A concerned El Niño item can be a forecast summary, an emergency warning, a request for preparedness, or a memory of past flooding. A misinformation flag can capture denial, conspiracy, unsupported physical explanations, false equivalence, or the subtler movement from a probabilistic forecast to a certain and dramatic claim. This is why the labels should not be read as self-explanatory counts. Their meaning depends on the communicative work being done in the post or comment.

5. Discussion

The central finding is that El Niño discourse is platformed. The same forecast-relevant issue appeared across six environments, but each environment made a different public visible. X/Twitter presented a high-volume, multilingual, fast-moving stream where weather, news, science, business, politics, agriculture, climate concern, and humor coexisted. YouTube presented a comment environment with higher negativity, anger, humor, skepticism, and personal experience. Facebook presented longer media/news-oriented posts with more calls to action and the highest model-flagged likely misinformation rate. Reddit presented a compact, weather-centered, concern-heavy discussion. TikTok presented short comments with more personal experience and expressive tone. LinkedIn presented a professional and informational version of the issue.

The possibility of automated or inauthentic activity does not remove the relevance of the platform differences, but it changes what they mean. If some visible items were produced or amplified by bots, fake profiles, organizational accounts, or coordinated actors, then the observed differences are differences among platformed information environments. This is still important for climate communication because users encounter a mixture of human, organizational, algorithmically amplified, and possibly automated content. However, it means that the results should be framed as differences in visible

discourse and platform-mediated attention and not as direct estimates of genuine public opinion.

Nonetheless, a single-platform study would produce plausible but partial conclusions. None of those conclusions would be necessarily wrong for the platform studied. The problem begins when the platform-specific conclusion is rewritten as a claim about social media, public opinion, or online discourse without cross-platform evidence. This point echoes prior warnings about representativeness and platform bias in social-data research [1,3]. It also extends the comparative logic that platform-specific affordances, publics, and genres are central to what researchers observe [7,25].

The examples also help explain why the platform differences are not only statistical. The same label can have different communicative meanings depending on the genre of the item. A skeptical item may be correcting exaggerated forecast language or advancing an unsupported causal theory. A personal-experience item may be a memory of earlier floods, a report of local weather, or a preference for warmer or cooler conditions. A humorous item may be playful, dismissive, anxious, or political at the same time. These differences fit the affordance-based framework when platforms shape not only how much people say, but what counts as a recognizable contribution [7,10].

For climate communication, the findings suggest that strategy should follow platform function. If the aim is to monitor fast-moving multilingual forecast discussion, X/Twitter is valuable in this corpus. If the aim is to understand personal experience, humor, confusion, and comment-level reaction, YouTube and TikTok deserve attention. If the aim is to monitor professional framing, LinkedIn reveals a different layer of discourse. If the aim is to observe concern within issue-specific communities, Reddit matters. If the aim is to watch public media/news posts and calls to action, Facebook matters. El Niño moves from forecast probability to public meaning as it passes through these environments. The same scientific signal can become a weather warning, a news item, a professional update, a joke, a skeptical comment, a food-security concern, or a memory of past storms.

The results suggest practical design principles for future cross-platform climate communication research. First, studies should define the intended inference before selecting platforms. Monitoring fast-moving forecast circulation, comment-level interpretation, professional framing, community concern, or misinformation risk may require different platform combinations. Second, researchers should preserve the distinction between communicative genres, for example, original posts, video comments, replies, subreddit threads, and professional updates, rather than treating all items as equivalent messages. Third, future designs should stratify samples by platform, language, and genre and should report platform-specific denominators instead of relying only on pooled totals. Fourth, when misinformation or stance labels are central to the argument, researchers should combine automated annotation with manual validation, repeated-call stability checks, and, where possible, bot or fake-profile screening.

For LLM-assisted methodology, this study shows both promise and caution. A shared annotation schema made it possible to compare heterogeneous platform traces using the same conceptual categories. That is valuable when the research question concerns external validity. At the same time, short comments, sarcasm, multilingual text, and ambiguous climate claims remain difficult measurement tasks. The current analysis should therefore be read as a structured computational comparison rather than a final coding truth. Future work should code a stratified validation sample by platform, language, and category, compare model labels with human judgments, estimate reliability, and repeat a subset of model calls to test stability [14–16].

6. Limitations

Several limitations shape interpretation of the findings. First, this study covers one issue during a nine-day event window. El Niño was selected because it was timely, visible, and scientifically defined, but other climate events, longer periods, or non-event-centered samples may produce different platform patterns. Second, retrieval was anchored on the keyword "El Niño", which may miss indirect discussion of heat, drought, flooding, agriculture, food prices, insurance, disaster preparedness, or weather risk that does not explicitly name El Niño. Third, collection is uneven because public platform access is uneven. This affects both the number of items and the type of items collected.

Fourth, the units of analysis are not identical. X/Twitter items are posts, YouTube and TikTok items are comments, Reddit is post-centered, and Facebook and LinkedIn contain public posts or comments where available. A comment under a video, a subReddit post, and a professional LinkedIn update are different genres. This study compares them because users encounter all of them as platformed discourse, but the comparisons should be read as differences among platformed communication environments rather than as perfectly interchangeable speech acts.

Fifth, this study does not measure the demographic diversity of users who posted or commented about El Niño. Platform publics may differ by age, gender, geography, profession, language, and other social characteristics, and the people who discuss El Niño on LinkedIn, TikTok, Reddit, YouTube, Facebook, or X/Twitter should not be assumed to be demographically interchangeable. Prior work has shown that demographic and gender-specific dimensions can affect social media discourse, sentiment, toxicity, and interpretation, while research on gender inference from social media also warns that inferring sensitive attributes from profiles, names, images, or text can introduce methodological error and ethical risk [26,27]. Given that the present dataset did not contain reliable, consent-based or platform-verified demographic information, this study does not infer users' age, gender, or location from names, profile images, language, or content. Future work should address user diversity through ethically collected metadata, opt-in surveys, data donation, platform-provided aggregate demographics, explicit self-descriptions where appropriate, or carefully validated demographic-inference methods and should test whether platform differences remain after stratifying by user characteristics.

Sixth, this study does not authenticate accounts or identify bots, fake profiles, organizational campaigns, or coordinated amplification. The findings therefore describe visible discourse in the collected public corpus, not verified human opinion. Automated or inauthentic accounts may contribute to platform differences, especially in volume, engagement, stance, and misinformation risk. Future work should combine text analysis with account-level metadata, posting histories, coordination indicators, and bot or fake-profile detection models where such data can be collected ethically and lawfully.

Sixth, the LLM annotation was not externally validated with a human-coded sample. The scripts used a shared schema, strict structured outputs, required fields, row matching, retries, and conservative prompt instructions, but these are pipeline safeguards rather than evidence that every substantive label is correct. LLM labels should therefore be interpreted as structured indicators for exploratory comparison. Future work should manually code a stratified validation sample across platforms, languages, and genres, report agreement metrics for key labels, examine disagreements, and repeat a subset of calls to assess stability.

Seventh, the instruction to use low-risk or unclear labels for ambiguous claims reduces false positives but may also reduce sensitivity to ambiguous skepticism, irony, coded misinformation, or humor. The likely misinformation estimates may therefore be conservative and should not be read as a definitive prevalence measure. Finally, ethical risks remain even when data are publicly accessible. This manuscript therefore reports

aggregate results, avoids handles and profile identifiers, does not reproduce searchable verbatim quotations, and treats raw records as sensitive research material.

7. Conclusions

This study began with a methodological question of how far a single-platform social media finding can travel. The El Niño case shows that the answer depends on the claim. Single-platform studies can be rigorous and valuable when they define their scope carefully. They become fragile when they describe one platform and then generalize to social media, public discourse, or public opinion more broadly. Across X/Twitter, YouTube, Facebook, Reddit, TikTok, and LinkedIn, El Niño discourse differed in volume, timing, language, topic, sentiment, emotion, stance, misinformation risk, personal experience, humor, calls to action, engagement, and length. The differences changed the story a researcher would tell.

The practical lesson is that platform choice is a research design decision with consequences for inference. Researchers should either keep conclusions platform-specific or design cross-platform comparisons when the intended claim concerns social media more generally. Climate communicators should make the same adjustment. The public around a forecast is not one public. It is a set of platformed publics, each with its own rhythms, languages, genres, and vulnerabilities. El Niño can be a forecast, a warning, a news story, a professional update, a joke, a skeptical comment, and a personal weather experience at the same time. Good research should make that plurality visible rather than collapse it into a single platform's view.

Author Contributions: Conceptualization: D.E. and N.K.; methodology, software, data collection, data curation, formal analysis, validation, visualization, and writing—original draft preparation: D.E.; supervision, project oversight, and critical revision of the manuscript: N.K.; writing—review and editing: D.E. and N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the HORIZON-MISS-2024-CLIMA-01 SystR Project, grant number 101212761.

Institutional Review Board Statement: Ethical review and approval were not required for this study. The study was a non-interventional computational analysis of publicly accessible, collector-visible social media traces and did not involve recruitment, direct interaction with individuals, surveys, interviews, experiments, intervention, access to private communications, private-account data, patients, biological material, or clinical research. Under Regulation (EU) 2016/679, Article 4(1) defines personal data as information relating to an identified or identifiable natural person, while Recital 26 clarifies that anonymous information, including information used for statistical or research purposes, is outside the scope of the data-protection principles where individuals are not or are no longer identifiable. The study also followed the research-safeguard principles reflected in GDPR Article 89, including data minimization. The analysis and manuscript use only aggregate and anonymized findings. Usernames, handles, profile details, URLs, and searchable verbatim quotations are not reproduced. Raw social media records are not publicly shared. Therefore, formal Institutional Review Board approval was not required.

Informed Consent Statement: Formal Institutional Review Board approval was not required for this study. The manuscript reports only aggregate and anonymized platform-level findings. Usernames, handles, profile details, URLs, and searchable verbatim quotations are not reproduced. Raw social media records are treated as sensitive research material and are not publicly shared.

Data Availability Statement: The aggregated dataset and analysis outputs supporting the findings of this study are available from the corresponding author upon reasonable request. Raw social

media records are not publicly shared because they may contain user-generated content and platform-specific identifiers. Access can be considered for research verification under appropriate ethical and data-protection safeguards.

Acknowledgments: During the preparation of this study, the authors used the OpenAI API, GPT-5.4-mini, for the purposes of LLM-assisted annotation, tagging, and consistency checking of public social media text data according to a predefined structured schema. The authors reviewed and edited the outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tufekci, Z. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the International AAAI Conference on Web and Social Media*; PKP Publishing: Burnaby, BC, Canada, 2014; Volume 8, pp. 505–514. https://doi.org/10.1609/icwsm.v8i1.14517.
2. Ruths, D.; Pfeffer, J. Social media for large studies of behavior. *Science* **2014**, *346*, 1063–1064. https://doi.org/10.1126/science.346.6213.1063.
3. Olteanu, A.; Castillo, C.; Diaz, F.; Kıcıman, E. Social data: Biases, methodological pitfalls, and ethical boundaries. *Front. Big Data* **2019**, *2*, 13. https://doi.org/10.3389/fdata.2019.00013.
4. Climate Prediction Center. *ENSO Diagnostic Discussion*; National Weather Service: Silver Spring, MD, USA, 2026. Available online: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/enso_advisory/ensodisc.shtml (accessed on 7 June 2026).
5. WMO. El Niño/La Niña Update (May 2026). 2026. Available online: https://wmo.int/resources/publication-series/el-ninola-nina-updates/el-ninola-nina-update-may-2026 (accessed on 7 June 2026).
6. FAO. El Niño. Retrieved 6 June 2026. Available online: https://www.fao.org/el-nino/en (accessed on 7 June 2026).
7. Matassi, M.; Boczkowski, P. An agenda for comparative social media studies: The value of understanding practices from cross-national, cross-media, and cross-platform perspectives. *Int. J. Commun.* **2021**, *15*, 207–228. Available online: https://ijoc.org/index.php/ijoc/article/view/15042 (accessed on 7 June 2026).
8. Hase, V.; Boczek, K.; Scharkow, M. Adapting to affordances and audiences? A cross-platform, multi-modal analysis of the platformization of news on Facebook, Instagram, TikTok, and Twitter. *Digit. J.* **2023**, *11*, 1499–1520. https://doi.org/10.1080/21670811.2022.2128389.
9. Oz, M.; Shahin, S.; Greeves, S.B. Platform affordances and spiral of silence: How perceived differences between Facebook and Twitter influence opinion expression online. *Technol. Soc.* **2024**, *76*, 102431. https://doi.org/10.1016/j.techsoc.2023.102431.
10. Bucher, T.; Helmond, A. The affordances of social media platforms. In *The SAGE Handbook of Social Media*; Burgess, J., Marwick, A., Poell, T., Eds.; Sage: Thousand Oaks, CA, USA, 2018; pp. 233–253. https://doi.org/10.4135/9781473984066.n14.
11. Pearce, W.; Niederer, S.; Özkula, S.M.; Sánchez Querubín, N. The social media life of climate change: Platforms, publics, and future imaginaries. *Wiley Interdiscip. Rev. Clim. Change* **2019**, *10*, e569. https://doi.org/10.1002/wcc.569.
12. Herasimenka, A.; Wang, X.; Schroeder, R. A systematic review of effective measures to resist manipulative information about climate change on social media. *Climate* **2025**, *13*, 32. https://doi.org/10.3390/cli13020032.
13. Vivion, M.; Trottier, V.; Bouhêlier, È.; Goupil-Sormany, I.; Diallo, T. Misinformation about climate change and related environmental events on social media: Protocol for a scoping review. *JMIR Res. Protoc.* **2024**, *13*, e59345. https://doi.org/10.2196/59345.
14. Törnberg, P. Best practices for text annotation with large language models. *Sociologica* **2024**, *18*, 67–85. https://doi.org/10.6092/issn.1971-8853/19461.
15. Ziems, C.; Held, W.; Shaikh, O.; Chen, J.; Zhang, Z.; Yang, D. Can large language models transform computational social science? *Comput. Linguist.* **2024**, *50*, 237–291. https://doi.org/10.1162/coli_a_00502.
16. Fang, Q.; Garcia Bernardo, J.; van Kesteren, E.J. A methodological guide on using large language models for reproducible text annotation in the social sciences and humanities with Python and R. *arXiv* **2026**, arXiv:2604.09638. https://doi.org/10.48550/arXiv.2604.09638.
17. University of Malta. Research Ethics and Data Protection on the Use of Data from Social Media. 2023. Available online: https://www.um.edu.mt/media/um/docs/research/urec/Harvestingdatafromsocialmedia.pdf (accessed on 7 June 2026).
18. Panel on Research Ethics. *Does Research Using Social Media Platforms Require Research Ethics Board Review? Guidance in Applying TCPS 2 (2022)*; Government of Canada: Ottawa, ON, Canada, 2023. Available online: https://ethics.gc.ca/eng/reb-cer_social-sociaux.html?wbdisable=true (accessed on 7 June 2026).

19. LSE Research Ethics Review Board. *Using Data from the Internet and Social Media in Research: Ethics & Consent*; London School of Economics and Political Science: London, UK, 2026. Available online: https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-Using-internet-and-Social-media-data.pdf (accessed on 7 June 2026).
20. Urban, O. Is Web Scraping Legal? Yes, If You Know the Rules. Apify Blog. 2026. Available online: https://blog.apify.com/is-web-scraping-legal/ (accessed on 7 June 2026).
21. OpenAI. Introducing Structured Outputs in the API. 2024. Available online: https://openai.com/index/introducing-structured-outputs-in-the-api/ (accessed on 7 June 2026).
22. Tumma, C.; Reddy, R.A.; Yadav, R.; Salati, A. Robust jailbreak detection in large language models using behavioral fingerprinting. In *2025 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom)*; IEEE: New York, NY, USA, 2025; pp. 284–289. https://doi.org/10.1109/MoSICom67153.2025.11398297.
23. Shukla, P.K.; Veerasamy, B.D.; Alduaiji, N.; Addula, S.R.; Sharma, S.; Shukla, P.K. Encoder only attention-guided transformer framework for accurate and explainable social media fake profile detection. *Peer-Peer Netw. Appl.* **2025**, *18*, 232. https://doi.org/10.1007/s12083-025-02047-z.
24. Daume, S.; Galaz, V.; Bjersér, P. Automated framing of climate change? The role of social bots in the Twitter climate change discourse during the 2019/2020 Australia bushfires. *Soc. Media + Soc.* **2023**, *9*, 20563051231168370. https://doi.org/10.1177/20563051231168370.
25. Hemphill, L.; Schöpke-Gonzalez, A.; Panda, A. Comparative sensitivity of social media data and their acceptable use in research. *Sci. Data* **2022**, *9*, 643. https://doi.org/10.1038/s41597-022-01773-w.
26. Thakur, N.; Cui, S.; Khanna, K.; Knieling, V.; Duggal, Y.N.; Shao, M. Investigation of the gender-specific discourse about online learning during COVID-19 on Twitter using sentiment analysis, subjectivity analysis, and toxicity analysis. *Computers* **2023**, *12*, 221. https://doi.org/10.3390/computers12110221.
27. Fosch-Villaronga, E.; Poulsen, A.; Søraa, R.A.; Custers, B.H. A little bird told me your gender: Gender inferences in social media. *Inf. Process. Manag.* **2021**, *58*, 102541. https://doi.org/10.1016/j.ipm.2021.102541.