

AI as We Describe It: How Large Language Models and Their Applications in Health are Represented Across Channels of Public Discourse

Jiawei Zhou
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
j.zhou@gatech.edu

Lei Zhang
College of Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
lzhang793@gatech.edu

Mei Li
College of Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
mli783@gatech.edu

Benjamin D Horne
School of Information Sciences
University of Tennessee Knoxville
Knoxville, Tennessee, USA
bhorne6@utk.edu

Munmun De Choudhury
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
munmund@gatech.edu

Abstract

Representation shapes public attitudes and behaviors. With the recent advances and rapid adoption of LLMs, the way these systems are introduced will negotiate societal expectations for their role in high-stakes domains like health. Yet it remains unclear whether current narratives present a balanced view. We analyzed five prominent discourse channels (news, research press, YouTube, TikTok, and Reddit) over a two-year period on lexical style, informational content, and symbolic representation. Discussions were generally positive and episodic, with positivity increasing over time. Risk communication was unthorough and often reduced to information quality incidents, while explanations of LLMs' generative nature were rare. Compared with professional outlets, TikTok and Reddit highlighted wellbeing applications and showed greater variations in tone and anthropomorphism but little attention to risks. We discuss implications for public discourse as a diagnostic tool in identifying literacy and governance gaps, and for communication and design strategies to support more informed LLM engagement.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → **Consumer health**.

Keywords

LLM, discourse, news, social media, science communication, public and professional narratives, public communication, AI literacy

ACM Reference Format:

Jiawei Zhou, Lei Zhang, Mei Li, Benjamin D Horne, and Munmun De Choudhury. 2026. AI as We Describe It: How Large Language Models and Their Applications in Health are Represented Across Channels of Public Discourse.



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/26/04
<https://doi.org/10.1145/3772318.3790316>

In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3772318.3790316>

1 Introduction

It is our use of a pile of bricks and mortar which makes it a 'house'; and what we feel, think or say about it that makes a 'house' a 'home'. In part, we give things meaning by how we *represent* them — the words we use about them, the stories we tell about them, the images of them we produce, the emotions we associate with them, the ways we classify and conceptualize them, the values we place on them. — Stuart Hall [36]

Large Language Models (LLMs) have rapidly captured public interest and have been embedded in many everyday information and communication systems like search engines [72, 80, 97]. In the health domain, LLMs are being explored for tasks ranging from communication assistance [50, 53] to decision support [103, 104]. However, evidence of these models' significant risks also has emerged, such as hallucinations and misinformation [17, 114, 120], algorithmic bias [49, 66, 75], and the potential to erode critical thinking and informed decision-making [46, 61, 101]. As such, (inter)governmental organizations and unions like the World Health Organization and the European Union have urged careful evaluation and regulatory governance [30, 88].

A core reason these issues are so complex is that LLMs differ fundamentally from prior information and communication technologies. Unlike traditional tools designed primarily to retrieve or organize existing information, LLMs generate new content in a probabilistic manner [106], popularly known as "next token predictors" [8]. This generative nature can create outputs that appear authoritative yet lack verifiable grounding or consistency [114, 120]. However, due to lagging public literacy and regulatory oversight, people's interactions are guided by pre-existing mental models and expectations shaped by prior technologies or even human relationships, including tendencies to anthropomorphize LLMs as intentional or emotionally aware [40]. As a result, users end up experimenting with this new technology for various reasons, often

in ways that are no longer appropriate or sufficient for understanding and evaluating LLM outputs [118, 119]. When people perceive LLMs as human-like or even human replacements, these flawed understandings can lead to real-world harms, as seen by the emerging reports of LLM-infused systems being linked to criminal or suicidal cases [26, 60, 111] and contemporary discussions of “AI psychosis” as a new type or force of mental issues [60].

Given these technical shifts and the potential misalignment between how LLMs work and how people expect them to behave, public discourse’s agenda-setting [69] function now plays an especially important role, as it “may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think *about*” [21]. With the rapid and wide adoption of LLMs, people’s understanding of these tools’ capabilities and risks will determine if they can mindfully and meaningfully decide how to engage with them. Past work has shown that media not only shape public perceptions of the salience of societal issues [69], but also influence behaviors and evaluations of emerging technologies and their potential impacts [23].

In other words, the way we introduce LLMs as an emerging technology will influence public perception and set expectations. Therefore, public discussions serve as both a proxy for public attitudes and interests in LLMs and as a force guiding future perceptions. However, less explored is whether current narratives offer a balanced view of both the potential and the limitations of LLMs.

We respond to this gap by analyzing and describing the public communication of LLMs in health, a high-stakes domain that touches everyone’s daily life. According to the agenda-setting theory [69, 70], there are two levels of agenda-setting in messages: the first-level agenda-setting is achieved through the “*salience of issues*” that affects public exposure and attention, and the second-level agenda-setting is delivered through the “*attributes of issues*” that influence public comprehension and sensation. Drawing on this, we examine both the *salience* of LLM-related issues by analyzing which information elements are communicated in public discourse, as well as the *attributes* of those issues by assessing the overall stylistic presentation of messages and the symbolic representation of message subjects (i.e., LLM entities). Specifically, this large-scale quantitative analysis characterizes how public discourse introduces the potential uses and risks of LLMs in the health domain across different channels by answering the following research questions:

- RQ1. [Lexical style]** What lexical tone and styles characterize public discourse on LLMs?
- RQ2. [Informational content]** What information on narrative framing, health subdomains, and risk disclosure is communicated in public discourse on LLMs?
- RQ3. [Symbolic representation]** To what extent are LLM entities anthropomorphically represented in public discourse?

To this end, we examined five prominent discourse channels between December 2022 and December 2024, based on the release date of ChatGPT (November 30, 2022) [2]. Each channel plays a unique role in shaping perceptions: news articles, research press releases, YouTube videos, TikTok videos, and Reddit posts. We collected 62,783 items using keyword combinations that required at least one health-related term and one LLM-related term, and reduced this to 21,773 items through LLM-assisted content filtering.

We then studied the general trends and cross-platform differences in lexical style, informational content, and symbolic representation. Our results demonstrate that public narratives about LLMs in health are generally positive and increasing so over time, and they are mostly framed episodically by focusing on isolated instances rather than broader societal or systemic implications. At the same time, we observe a lack of a thorough introduction to or an overview of LLM risks in public discourse. When risks are mentioned, they are largely framed as information quality concerns, with rare explanation of the generative nature that differentiates LLMs from traditional information sources. Further, differences exist across content creator types: layperson-driven platforms such as TikTok and Reddit emphasize consumer needs and mental health care while showing greater variations in emotional tone and anthropomorphism, yet are less likely to communicate risks. On the contrary, professionally-authored content such as news articles and research press releases focus more on clinical applications and are more likely to situate issues within broader societal or systemic contexts.

Overall, this work makes three key contributions. (1) We offer a large-scale and comprehensive examination of public discourse across five different communication channels. To the best of our knowledge, this is the first study to study discourse across both top-down (professionally created) and bottom-up (publicly generated) content on LLMs and their applications in health. (2) We categorize six health subdomains where LLMs can be applied, which can guide future research on understanding and evaluating LLM deployment in health contexts. (3) We provide empirical evidence on the current state of public discussions about LLMs and identify literacy and governance gaps. We discuss the implications for using discourse as a diagnostic tool to understand public perception and attention and for developing more effective communication approaches and design strategies to support user agency and knowledge for meaningful and mindful engagement with LLMs.

2 Related Work

2.1 Media Effects on Public Perceptions of Emerging Technology

Media, ranging from mass media to social media, has long been recognized for its role in both reflecting and shaping public perceptions and attitudes [15, 69, 100]. For instance, media exposure has been shown to drive public engagement in national conversations [58], influence views on the importance and priorities of regulatory issues [31, 55, 69], and can both inform [27, 37] and misguide [28] the public.

With regards to emerging technology, media plays a key role in communicating the sociotechnical risks and benefits [9, 67, 109]. Public discourse coverage can affect which aspects of a technology are emphasized, which in turn shapes how the public evaluates its value and impact. For example, public discourse can emphasize or de-emphasize who is responsible for benefits and harms of a technology, specific use cases of the technology, or the need for regulation [109]. As a result, evidence from studies and opinion polls suggests that public attitudes towards emerging technology can “mirror the news media’s stance” [6, 67]. Specifically, prior work has studied media coverage of emerging technologies such as nanotechnology [4] and AI [3, 44, 89], including relevant

risks [3, 109] and technological issues [24]. Broadly, these works have demonstrated that media tends to embrace a positive outlook about new technology, focusing more on the benefits than the potential risks [4, 112], although with some technologies this trend has changed over time [109]. In the case of AI, Ouchchy et al. [89] found that media portrayals of AI ethics issues were shallow and largely focused on practical implications.

2.2 Public Attitudes Toward LLMs and Generative AI

Generative AI and specifically large language models (LLMs) are unique as their use in consumer-facing products has been quickly widespread [98], letting the public experiment with the technology themselves ahead of adequate public literacy and regulatory oversight. This public experience, alongside the rapid technological development, likely changes both how influential the public discourse is in perception formation and how the public discourse covers the technology. As such, public narratives surrounding LLMs and LLM-infused systems may be more participatory and reflective of personal experiences than those surrounding prior emerging technologies like nanotechnology.

Recent work has begun to explore public perceptions of AI-generated content [52, 65] and perceptions of AI itself [20, 78]. However, overall findings remain mixed, and public attitudes were found to vary across socio-demographic factors [78, 96]. Some work points to skepticism toward AI-generated content. Kadoma et al. [52] studied the perceptual harms of generative AI and found that people tended to associate AI writing with lower quality. Other studies have also shown negative perceptions about AI-generated content. For example, Wang et al. [107] found that people felt negatively about AI-generated art on TikTok due to both qualities of the content itself (e.g., hyper-realistic or unsettling qualities of the paintings) and principled stands against AI in art (e.g., “unfairly copying or appropriating the work of human artists”). On the other hand, other research indicates favorable perceptions. Park et al. [90] discovered that people perceived AI Instagram accounts as more attractive and the quality of their content as similar as influencer-created content. More broadly, Cheng et al. [20] found that Americans generally viewed AI as warm and competent, and this perception has significantly increased over time.

Overall, current work suggests that public attitudes toward AI are neither uniformly positive nor negative, but instead context-dependent. Evidence shows that both AI usage and attitudes are mixed and vary across industries [105] and across populations [78, 96]. For instance, a global study of college students found that students generally felt positive about using ChatGPT, but its use varied across socio-demographic and geographic factors [96].

While these studies offer valuable insights into how the public perceives LLMs and LLM-generated content, they largely focus on specific platforms and contexts. There remains a gap in holistically understanding how LLMs are represented across diverse discourse channels, including both professional-led and public-driven sources. Our study builds on prior scholarship by offering a large-scale examination of both top-down (professionally created) and bottom-up (publicly generated) content.

2.3 LLMs and Generative AI in Health and Healthcare

Besides the fact that people are experiencing LLMs before adequate understanding and oversight, LLMs are also distinctive due to the breadth of proposed use cases, ranging from general-purpose ones such as LLM-infused search engines [84] to specific ones like LLM-infused health record systems [59]. In this study, we focus on LLMs within the high-stakes domain of health that influences people’s daily life. In health, LLMs have been shown to have potential in supporting a variety of tasks such as clinical documentation [103, 117], medical question answering and reasoning [5, 82, 117], clinical decision support [103, 104], therapeutic conversations [57, 99, 108], and public health interventions [50, 51, 53]. In practice, LLMs have already been implemented in existing systems or workflows, including Electronic Health Records (EHR) systems [59], virtual agents for dealing with domestic violence [68], and public health interventions [50]. Further, public-facing LLM chatbots like ChatGPT are being used for seeking health information [116].

Evidence so far suggests that perceptions of generative AI in healthcare have been mixed but skew positive, similar to perceptions of generative AI in broader contexts. A study by Heinrichs et al. [39] showed that AI in medicine is viewed positively, with physicians reporting high AI enthusiasm and low AI skepticism. On the other hand, though a randomized experiment, Yang et al. [115] demonstrated that practicing clinicians perceived physicians using generative AI as a primary decision-making tool as having “significantly lower in clinical skill.” Still, the same study supported general enthusiasm about generative AI’s use in healthcare. Others have similarly suggested that patients have a positive view of generative AI in health. For instance, through a large global survey, Busch et al. [12] found that the majority of patients held positive views about AI in healthcare. However, once again, perceptions varied widely across socio-demographic groups and technical literacy.

2.4 Risks of LLMs and Generative AI

Despite the largely positive view of generative AI among the public, researchers have raised concerns about the limitations and risks of adopting LLMs in the real world. General issues include representativeness of training data [113], privacy and security [118], language and geographic disparities [49, 75], the generation of inaccurate, biased, or toxic content [17, 62, 120], and the meaningfulness of evaluation metrics [113]. Because of these limitations, questions remain regarding LLMs’ ability to meet clinical standards [38], avoid race-based medical misconceptions [85], attend to emotional needs [99, 108], or handle high-stakes or high-intensity cases [57, 99]. Such pitfalls are not simply technical challenges; they may contribute to serious real-world harms, as seen by the emerging reports of LLM-infused systems being linked to criminal or suicidal cases [26, 60, 111].

Research has shown that people tend to struggle to recognize technological changes and instead rely on perceived reliability [76]. In the case of LLMs, work has suggested that people tend to hold diverse and sometimes flawed mental models of LLM systems, which are relevant to AI knowledge. Evidence has showed that people tended to have overly simplified and erroneous mental models of data handling and model mechanisms, which limited risk awareness

and judgment [118]. People without basic AI literacy skills struggle more in critically engaging with LLMs [93]. Regarding these gaps in understanding and attitudes, research suggests that they are relevant to AI literacy and societal framing. Studies have identified polarization in AI attitudes linked to differences in technological knowledge, with misalignment between AI influencers and the U.S. general population [78] and especially strong polarization in tech-centric communities [94].

What is worse, gaps in risk perceptions and AI literacy can be amplified by society's imbalanced representations of LLMs' capabilities and limitations, potentially leading to societal hype. Drott et al. [25] criticized claims of AI outperforming medical practitioners, as much of the supporting evidence is not empirically convincing or transparently reported. Global news media on AI risks also had a skewed focus on societal, legal and rights-related risks, which may leave the public with an incomplete understanding of potential harms [3]. However, despite these findings, less is known about how LLMs are framed and symbolically represented across channels of discourse, leaving open questions about how conceptualizations of a novel technology may shape public understanding and expectations. We respond to this gap by considering how societal issues and AI's identity are framed in public discussions.

3 Method

This study presents a large-scale, multi-platform analysis of how large language models (LLMs) and their applications in health are represented in public discourse. As shown in Fig. 1, we examined five prominent discourse channels: news articles, research press releases, YouTube videos, TikTok videos, and Reddit posts — each playing a distinct role in shaping public understanding and attention. Using health- and AI-related keywords, we first collected data across these platforms and applied an additional round of LLM-assisted filtering to ensure content relevance to LLM adoption in health fields. Our analysis focuses on three key dimensions: (1) *Lexical style*, which captures overall presentation style through emotional tone and writing style; (2) *Informational content*, which analyzes the mentioning of specific information surrounding framing, health subdomains, risk and mechanism disclosure; and (3) *Symbolic representation*, which examines the framing of the AI's identity through the level of anthropomorphism attributed to LLM entities. To assess differences across channels, we employed Kruskal–Wallis H tests to compare distributions across groups for each dimension, followed by Dunn's post hoc tests with Bonferroni correction for pairwise comparisons.

3.1 Data

We studied five distinct media sources: news articles, research press releases, YouTube videos, TikTok videos, and Reddit posts. Each source plays a unique role in shaping perceptions. To elaborate, news coverage by journalists aims to make complex issues accessible to general audiences but may oversimplify nuances or overemphasize aspects that are more likely to resonate with the public. In contrast, research press releases are official statements produced by scientific organizations, such as universities and medical centers, to share research findings with the general public. They represent

more in-depth and expert-led coverage but tend to be less accessible and have a narrower focus. YouTube, as a platform for video content, serves as a space for both professional and user-generated content, where discussions about LLMs can vary widely in terms of depth and perspective. TikTok, with its younger user base [77], presents a more informal, quick-hit style of content, and it may prioritize entertainment over depth and objectiveness. Lastly, Reddit's community-driven structure and pseudo-anonymous design foster conversations among individuals with shared interests and encourage more focused discussions about emerging technologies and personal experiences with those technologies [94].

Ethics Statement: As we used publicly accessible data without any direct interactions with individuals, this work did not require institutional review board approval. However, we took careful attention in managing and presenting our data and followed the ethical practices in social media research by removing deleted or removed content and paraphrasing social media content quoted as examples.

3.1.1 Data Source and Collection: We collected public narratives in English over a two-year period, from December 2022 to December 2024, based on the release date of ChatGPT (November 30, 2022) [2]. To curate related content discussing LLM for health, we used a combination of health- and LLM-related keywords with an 'OR' operator within a category and an 'AND' operator across, ensuring each item contains at least one LLM-relevant keyword (e.g., “*language model*” or “*ChatGPT*”) and at least one health and symptom term (e.g., “*well-being*”, “*drug*”, or “*tingling*”). The complete keyword list can be found in Appendix A. These keyword was adapted from search strategies in existing review studies and WHO-listed health topic words [64, 86, 91]. For instance, Patel et al. [91] used “*Generative AI/GenAI/Large Language Models*” as the search terms for systematic reviews. After several rounds of sanity checks by the authors, keywords were refined by removing ones that over-introduced false positives. For example, ‘gemini’ was replaced with ‘google gemini’ to avoid confusion with the word ‘gemini’ in astronomy. In total, we collected 62,783 items of articles, posts, and videos. The data size for each source is presented in Table 1. Specifically, data from each discourse channel was collected as follows:

- **News:** News articles were extracted from extended versions of the NELA-GT [35] and NELA-Local [41] datasets using the same collection pipelines, which include articles published by U.S. national and local news media outlets. We limited media outlets to those deemed reliable mainstream sources or mixed reliable sources, based on the Media Bias Fact Check (MBFC) [18], an independent resource maintained by researchers and journalists to evaluate the bias and credibility of media sources. The goal is to exclude fringe news without standard editorial oversight.
- **Research Press:** Research press release data was collected from the EurekAlert [33] news-release distribution operated by the American Association for the Advancement of Science. EurekAlert hosts press releases produced by organizations that engage in all disciplines of scientific research, such as universities, medical centers, and government agencies.
- **YouTube and TikTok:** For video data, we first identified relevant videos in English using *YouTube Data API* and *TikTok Research API*, and then collected and transcribed identified videos

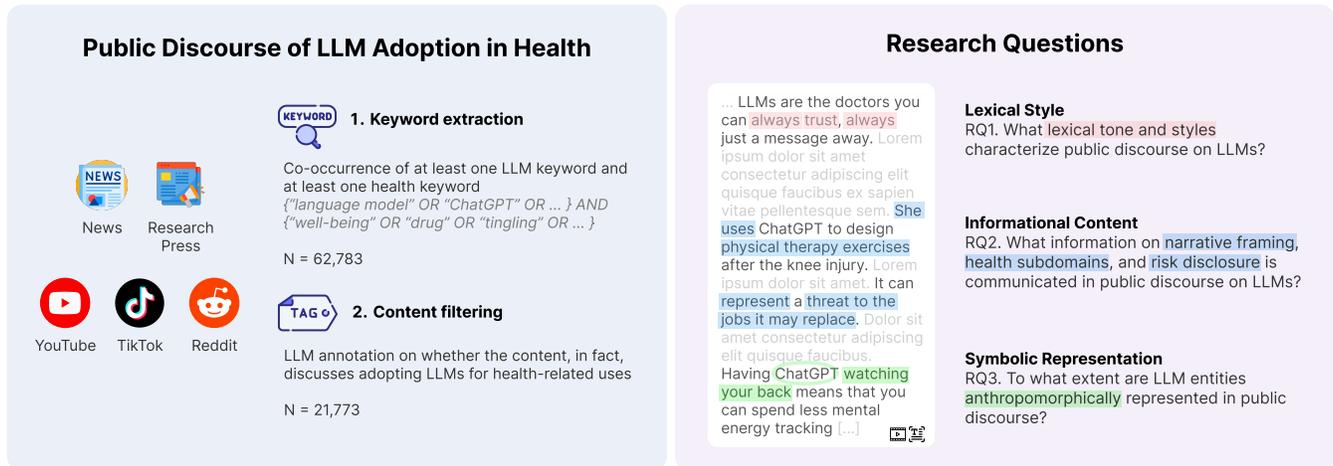


Figure 1: Overview of our study investigating how public discourse introduces LLMs and their applications in the health domain. We collected public narratives from five prominent discourse channels, including news, research press, YouTube, TikTok, and Reddit, between December 2022 and December 2024. Drawing on agenda-setting theory, we examined three core dimensions of discourse: (1) Lexical Style: the overall presentation style of discourse through emotional tone and writing formality; (2) Informational Content: how messages frame LLMs’ implications, risks, mechanisms, and potential across health domains; (3) Symbolic Representation: the level of anthropomorphic representation of LLM entities.

for subsequent analyses. We decided not to rely on the transcripts available through *YouTube Transcript API* and *TikTok Research API* due to inconsistent transcription quality on YouTube API and the inconsistent availability of built-in voice-to-text features on TikTok API. Instead, we referred to prior work [81] and implemented a video transcription pipeline using *yt-dlp* to download videos, *ffmpeg* to extract audio, and *whisper* [95] to transcribe the audio content. This approach allowed us to collect transcriptions for the majority of videos, except for 693 videos consisting of pure music or no audio, 158 videos that were inaccessible due to privacy restrictions, 169 videos with encoding issue that prevented audio extraction, and 155 videos with gibberish content. Additionally, an internal TikTok API error¹ prevented data collection starting from August 18, 2024. We had submitted a support ticket to TikTok, but the API internal issue was not resolved at the time of writing and revision.

- **Reddit:** Reddit data was extracted from the widely used PushShift dataset [7], with deleted and removed content excluded from analyses. We only included the original submissions, not the replies, to ensure sufficient context for analyses.

3.1.2 Content Filtering: While the above crafted keyword combination ensures the extracted content mentions words or phrases about both LLMs and health, it is possible that such expressions appear in unrelated contexts, such as “people *abusing ChatGPT* for writing their essays.” Therefore, we conducted a round of targeted filtering to ensure that the content, in fact, discussed adopting LLM for health-related uses. We used GPT-4o-mini [43] with few-shot examples to assist this additional round of content filtering. We

¹Similar issue has been reported online: e.g., <https://stackoverflow.com/questions/78731820/how-to-resolve-error-code-500-for-tiktok-research-api>; https://www.reddit.com/r/CompSocial/comments/1d8mdz5/tiktok_api/

Table 1: Descriptive summary of the number of items and word length across the five channels of public discourse.

Source	Items		Length (# words)		
	(pre filter)	(post filter)	Mean	Median	Std dev
News	6 550	2 254	1 158.74	898	1 247.01
Research Press	867	634	805.96	742	411.14
YouTube	6 632	3 559	5 451.65	3 228	8 958.01
TikTok	9 521	3 272	276.07	228	234.29
Reddit	39 213	12 054	510.81	297	697.64
	62 783	21 773			

referred to previous work [74] and set the temperature as 0.0, provided four examples (3 positive examples and 1 negative example), and emphasized that annotations are based solely on the literal writing of the text. The complete prompt is provided in the supplement. For cases where the content exceeded the 128k token limit, the articles were split into several chunks, and the annotation results were combined. After this targeted filtering, the final dataset was reduced from 62,783 items to 21,773 items. Table 1 provides a summary of item numbers and content length statistics across the five sources. To assess the reliability of utilizing an LLM for filtering articles, we sampled 100 articles across five data sources for human annotation. The model performed well on this task with an F1-macro score of 0.9 and Cohen’s κ of 0.7947.

3.2 Lexical Style

To understand the overall presentation style in different discourse channels portraying LLMs, we analyzed both emotional tone and writing style, as these summary patterns would better capture the overall communication styles than other direct word-category

counts. We analyzed emotional tone to assess the pattern of affective expressions, to understand how different discourse channels portray LLMs in emotionally charged or neutral ways. For this, we utilized the Linguistic Inquiry and Word Count (LIWC) [102] program, which is a validated psycholinguistic lexicon and has been widely used in linguistic and social science research to study emotional and cognitive processes in text [48, 120]. *Emotional tone* is quantified on a continuous scale, with values close to 0 indicating a highly negative sentiment, 50 representing a neutral sentiment, and values approaching 100 corresponding to a highly positive sentiment. *Analytic style* (or categorial-dynamic index, CDI [92]) captures expressions of abstract thinking and cognitive complexity, which are associated with formal and logical thinking patterns. *Clout style* reflects the prevalence of self-focused expressions that indicate relative social status, confidence, and leadership, which tends to capture perceived authority, confidence, and credibility and is commonly studied in health communication and discourse [13, 53]. Lastly, *authentic style* indicates the level of spontaneous and non-regulated language and is commonly used in self-disclosures.

3.3 Informational Content

Following our analysis of writing style, we analyzed three features about the informational content: what *health subdomains* are mentioned, how each discourse channel *frames* LLMs in health, and if and how *risks* and LLM mechanisms are disclosed in the content.

3.3.1 Framing Types and Dimensions. We analyzed how content is framed by examining 1) *framing type* – whether the narrative is framed episodically or thematically, and 2) *framing dimension* – what broader policy-related themes are framed as relevant. These two framing typologies were proposed by Iyengar [45] and Boydston et al. [10] respectively. *Framing type* focuses on how the story is told, distinguishing between episodic and thematic approaches. Specifically, episodic framing presents concrete information about specific people, places, or events, often focusing on individual stories or case studies. On the other hand, thematic framing provides a more generalized perspective on issues, placing stories in broader political and social contexts. *Framing dimension* identifies frame cues across 14 policy issues, categorizing the content within broader societal and political contexts such as health, technology, or societal impact. To examine these framings, we used labels for both narrative and issue framing from a dataset of 4500 manually annotated data [71]. Then, for each type of frame, we built a multilabel RoBERTa classification model, replicating prior work [71] to classify the frames in all items across the five discourse channels. This model has been used in prior work to study gender violence in news [73]. Using both the original definitions by Boydston et al. [10] and Iyengar [45] and the comprehensive codebook by Mendelsohn et al. [71], we annotated 100 items to evaluate the applicability of this model on our dataset and found the model performed with an F1-macro average score of 0.7761 (Precision: 0.8, Recall: 0.7536).

3.3.2 Risk and Generative Nature Disclosure. We considered whether the content discussed the potential risks of LLMs and the differences in mechanisms from traditional information technology. To assess these aspects, we utilized a risk taxonomy focused on LLM adoption for public health [119]. This typology was selected for its relevance

to the health domain and its greater applicability compared to other existing taxonomies. Based on focus groups with health professionals and individuals with experience in seeking health information, this work identified distinctive characteristics of LLMs that differentiate them from traditional information sources and summarized potential negative consequences of adopting LLMs for health informational needs [119]. The taxonomy identifies four risk dimensions: (1) risks to individuals: how LLMs may negatively influence individuals' actions, decisions, and overall well-being, (2) risks to human-centered care: how LLMs may undermine the relations and systems of healthcare and social support network, (3) risks to the information ecosystem: how LLMs may affect the people, technologies, and norms that shape the way information is produced, shared, and evaluated, (4) and risks to technology accountability: how LLMs may challenge existing mechanisms of oversight, regulation, and security in technologies. In addition to the risk categories, we also explored whether the content explained the *generative nature* of LLMs (i.e., mentioning of probabilistic nature and limitations such as no real understanding of content) compared to non-generative AI or information retrieval-based systems [8, 106, 119].

We utilized GPT-4o-mini [43] with few-shot examples to annotate the presence or absence of each risk category and its generative explanation. This approach allowed us to efficiently scale the annotation process across the corpus. Consistent with previous task settings described above, we set the temperature to 0.0 to ensure deterministic output and provided four examples to guide the model in making accurate classifications. To assess the reliability of utilizing LLM for categorization, we randomly sampled 100 articles across five data sources for human annotation, where risk disclosure had an F1-macro score of 0.74 (Precision: 0.72, Recall: 0.79), and generative nature explanation had an F1-macro score of 0.85 (Precision: 0.95, Recall: 0.80). To further examine potential model-specific bias, we conducted an additional cross-model validation using gemini-2.5 on a random subset of 500 items. gpt-4o-mini showed slightly higher F1-macro scores with good inter-model agreements, indicating stable annotation behavior on our data. We report both the model-human evaluation (n = 100) and model-model agreement (n = 500) statistics in Appendix 2.

3.3.3 Health Subdomain Prevalence. The potential applications of LLMs in health span a wide range, from providing healthy lifestyle advice to supporting mental health and professional medical training. Therefore, to understand the distribution and prevalence of specific health subdomains mentioned in the content, we combined human annotation with the use of GPT-4o-mini [43] to scale up the categorization across the entire corpus. Specifically, we began by conducting conventional content analysis [42] to summarize the health subdomains referenced in the content, based on a random sample of 200 items that were equally distributed across the five discourse channels. Two researchers independently reviewed all the sampled items to establish a general understanding of the data. Afterwards, they used an inductive and iterative approach to develop new codes or categorize each instance into existing codes. In the meantime, the codes were continuously revisited for potential modification and refinement, either by breaking them into sub-categories or by grouping them into broader categories. By the end of the analysis, we determined that the categorization had reached

saturation without emerging meaningful subdomains. This subdomain categorization was reviewed by two experts: a practicing medical doctor with research experience and a researcher specializing in public health communication. This expert consultation helped us validate and improve the clarity of all categories.

The final categorization comprised 20 subcategories, organized into six broad groups: clinical decision support, workflow optimization, professional training, research and data analysis, consumer health support, and healthcare system support. The complete set of subcategories is presented in the Results Sect 4.2. To scale up the categorization across the entire corpus for prevalence analysis, we employed GPT-4o-mini [43] using the same temperature settings of 0.0 as previous tasks to ensure deterministic output. The prompt is provided in the supplement. When the content and prompt together exceeded the token limit, the content was split into multiple chunks, and the annotation results were combined for final reporting. To assess the reliability of utilizing LLM for categorization, we sampled 100 articles across five data sources for human annotation. The model performed fine on this task with an F1-macro score of 0.792 (Precision: 0.745, Recall: 0.888). Same as previous tasks, we conducted an additional cross-model validation using gemini-2.5 on a random subset of 500 items. gpt-4o-mini showed slightly higher F1-macro scores than gemini-2.5 with high inter-model agreement, indicating stable annotation behavior on our data. We report both the model-human evaluation ($n = 100$) and model-model agreement ($n = 500$) statistics in Appendix 2.

3.4 Symbolic Representation

Lastly, to examine the symbolic representation of LLMs in public discourse, we measured the level of implicit anthropomorphism in expressions containing LLM entities. We used a masked language model to capture the degree of anthropomorphism associated with LLMs in a given sentence. This approach proposed by Cheng et al. [19] utilizes a masking technique to compare how much an entity is implicitly framed as human versus non-human, by computing the log of the ratio between the model's output probabilities for replacing the mask with human pronouns versus non-human pronouns. Specifically, for each instance in our corpus, when a sentence mentioned any LLM entity, the full sentence was extracted, with the LLM entity replaced with a <MASK> token. In addition to the list of LLM entities provided in the original work [19], we supplemented the model with our LLM-related keywords used in the content filtering step. An anthropomorphism probability was then calculated for each identified entity in a masked sentence (Equation 1, where *Anthro* captures the degree of anthropomorphism for entity x in sentence s), which was then used to calculate the average anthropomorphism score across all identified masked sentences within a given item. With this approach, an anthropomorphism score of 0 means that the LLM entity is equally likely to be stated as either human or non-human, and positive scores mean that the LLM entity is more likely to be implicitly framed as human in the sentence.

$$Anthro(s_x) = \log \frac{P_{Human}(s_x)}{P_{Non-Human}(s_x)}. \quad (1)$$

4 Results

4.1 Lexical Style

Summary: (1) Public discussions on LLM for health were generally positive, and this positivity has gradually increased over time. TikTok and YouTube showed the most optimistic tone, whereas the News had an average neutral tone. (2) Writing styles mirrored these patterns: News and Research Press had highly analytical styles, consistent with professional norms. In contrast, social media platforms were less analytical and TikTok and YouTube scored higher on clout, suggesting that content creators presented themselves as confident and persuasive. Across all channels, authentic and spontaneous communication was low, particularly in professional outlets.

4.1.1 Emotional Tone. Overall, the emotional tone of public discussions on LLM for health was generally positive, with most discourse channels (except for News) showing a positive emotional tone on average. This suggests that the portrayal of LLMs in health contexts was largely optimistic. In addition, as shown in Fig. 2, we found the median emotional tone in public discourse on LLMs had increased over time despite some fluctuations. However, there was notable variation across channels. TikTok and YouTube content tended to exhibit the most positive tone, with higher median and mean values. For example, one video says “*ChatGPT designed a **unique healthy** recipe that tasted **great** [...] I had one minor tweak, but overall **great** tool, **great** time-saver, very **exciting** new technology.*” Dunn’s post hoc test further demonstrated that YouTube and TikTok content was significantly more positive than other sources. In contrast, News and Reddit discussions showed a more neutral tone, with Reddit in particular having the widest range of emotional attitudes. Dunn’s post hoc test indicated that News was significantly more negative than other sources. This could be explained by the nature of these platforms: Reddit’s community-driven nature, which encourages diverse opinions and facilitates focused discussions, while news articles are more likely to reflect journalistic norms that emphasize balance between highlighting opportunities and addressing risks. The technological orientation of Reddit’s community may also have contributed to this high variation, as Qi et al. [94] has shown that tech-centric communities exhibit greater polarization.

4.1.2 Writing Style. We observed differences in writing style across discourse channels (Fig. 3), which reflects the underlying communication goals and norms of each venue. First, News and Research Press displayed the highest level of analytical writing that is related to logical reasoning and formal communication. This tendency aligns with the expectation of rationality and professionalism in institutional communication. In comparison, the other three user-driven social media platforms exhibited lower and more varied levels of analytical style, which suggests a tendency for these social media platforms to prioritize accessibility over formal reasoning.

The second dimension, clout, captures self-focused expressions that are associated with confidence and leadership. Contrary to analytical expressions, YouTube and TikTok had higher scores, indicating that content creators on these two platforms tended to present themselves as opinion-influencing figures. These two platforms tended to use more collective pronouns (e.g., “*we love*

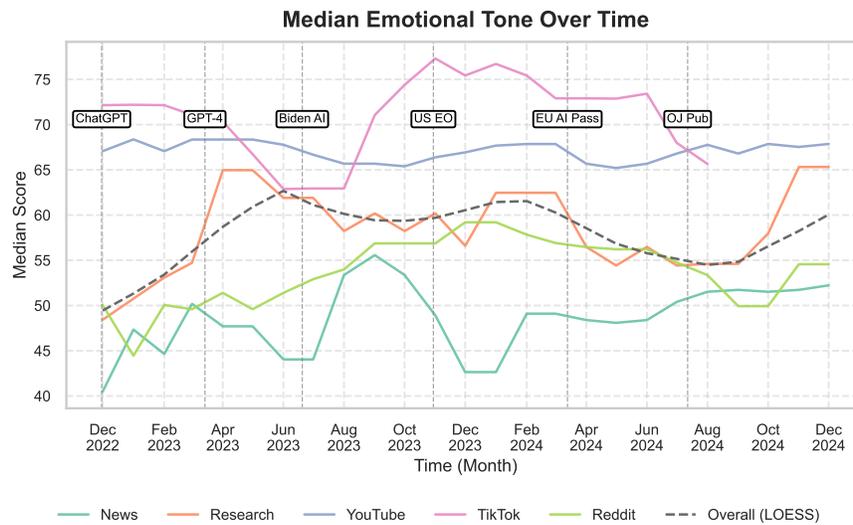


Figure 2: Trends in median emotional tone, where public discussions of LLMs in health contexts were largely positive (except for news) and became increasingly positive over time. TikTok data ended in August'24 due to an API internal error (see Sec. 3.1.1). LOESS (Locally Estimated Scatterplot Smoothing) was applied to capture overall trends across data sources of varying size.

that ChatGPT thinks like us”) or second-person singular pronouns (e.g., “so **you** have downloaded ChatGPT and played around with it. **you** liked it but **you**’re probably asking how do I save money with it”), potentially because these platforms encourage videos with one-to-one conversational styles and expressions of shared feelings. Dunn’s post hoc test further demonstrated that YouTube and TikTok had significantly higher clout scores than other channels. Both platforms also used more certainty words (e.g., *never, definitely*) and fewer tentative words (e.g., *maybe, perhaps*); for instance, “they are the doctors you can **always** trust, **always** just a message away.”

Lastly, all public channels incorporated a lower level of authentic communication styles, suggesting they had a high degree to which a person is self-monitoring. This lower authentic tendency suggests that cognitive process words could be used to support statements, and the overall language is more abstract. Among them, News and Research Press had the lowest average levels, which can be explained by expectations for professional writing to avoid personal or informal language.

4.2 Informational Content

Summary: (1) Public discussions of LLMs in health emphasized specific and isolated cases, with episodic framing more common than thematic framing. Practical implications for healthcare, safety, and the economy were highlighted more, while broader systemic or policy-related considerations were rarely touched on, except in the news. (2) Overall, public discourse lacked a thorough mention of risks, and the generative nature and distinctive affordances of LLMs were rarely discussed. When risks were discussed, information quality issues were most frequently highlighted. (3) Layperson-driven platforms were especially less likely to mention risks or explain generative nature than institutional sources, while YouTube served

as a middle ground with both organization- and user-generated content. (4) In terms of specific health subdomains, TikTok and Reddit emphasized consumer-facing applications, particularly mental health support, while News and Research Press highlighted clinical decision support and research-oriented applications.

4.2.1 Framing Type and Dimensions. For framing type, we found that episodic framing was noticeably more prevalent than thematic framing across all discourse channels (Fig. 4a). This means that LLM adoption in health is often introduced through specific and isolated cases, such as spotlighting particular use cases or technological milestones, rather than being presented as a generalized and systematic happening. Kruskal–Wallis H-tests showed significant differences in framing types across channels, where episodic $H = 715.82$ (***) and thematic $H = 1310.74$ (***). Notably, News showed a significantly higher proportion of thematic framing (31.5%) than other sources, which may be explained by journalistic writing patterns that tend to situate events within larger societal contexts.

In terms of framing dimensions, direct and practical impacts on healthcare, safety, and economy were more common, while broader systemic, ethical, or policy-related framings remain underrepresented, particularly in social media platforms (Fig. 4b). Unsurprisingly, given this study’s focus on LLMs in health, health and safety had a higher prevalence, ranging from 19.8% in YouTube to 54.3% in Research Press. The next most frequently mentioned social issue dimension was economic, especially for News (42.4%) and TikTok (26.9%). A closer examination of data showed some subtle differences: News tended to emphasize corporate and macroeconomic implications, while TikTok had more discussions on personal financial impacts. Nevertheless, disruption of the labor market appeared as a consistent theme, ranging from personal worries (e.g., “I think I’m officially out of a job. I asked ChatGPT to make a knee rehab

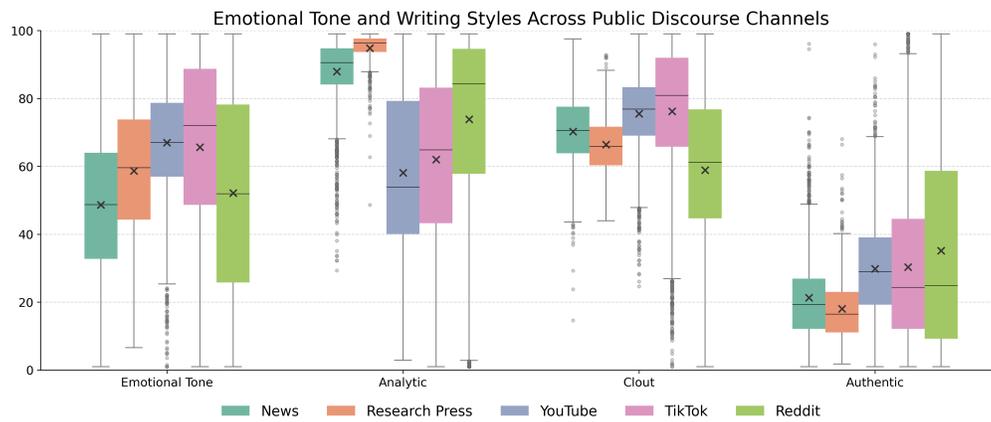
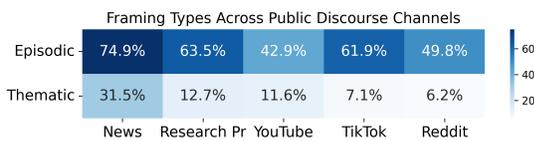
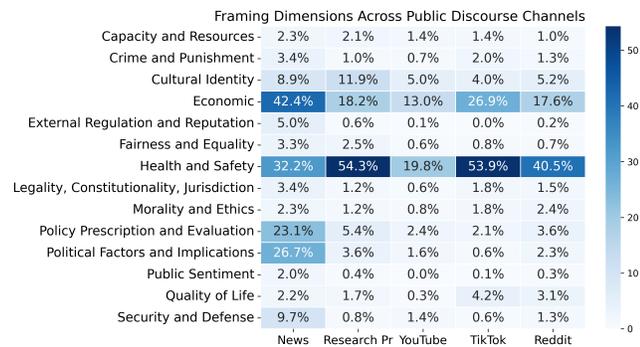


Figure 3: Lexical style in public discourse on LLMs for health. Kruskal–Wallis H-tests were performed to determine whether there was a significant difference across channels (** $p < 0.01$, * $p < 0.05$), where emotional tone $H = 1332.73$ (** $p < 0.001$), analytic $H = 3552.45$ (** $p < 0.001$), clout $H = 2895.86$ (** $p < 0.001$), and authentic $H = 419.56$ (** $p < 0.001$). Dunn’s post hoc test indicated that content on YouTube and TikTok was significantly more positive, while news was significantly more negative than other channels.



(a) Framing types ($p < 0.001$): whether narratives are framed episodically or thematically.**



(b) Framing dimensions ($p < 0.001$): what broader policy-related themes are framed as relevant.**

Figure 4: Framing types and dimensions in public discourse on LLMs for health. Kruskal–Wallis H-tests showed significant differences at all levels across channels (** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$), with higher H in thematic framing and discussions related to policy prescription and evaluation, as well as political factors and implications.

exercise plan, and it came up with this in seconds...” to broader societal impacts (“Although conversations surrounding technological unemployment over the past several decades have revolved around

blue-collar workers losing their positions to automated robotics solutions, the widespread use of ChatGPT has introduced similar questions in the healthcare and other knowledge-based professions.”).

Kruskal–Wallis H-tests showed significant differences in framing dimensions across channels, where capacity and resources $H = 28.44$ (** $p < 0.01$), crime and punishment $H = 74.14$ (** $p < 0.001$), cultural identity $H = 102.03$ (** $p < 0.001$), economic $H = 863.32$ (** $p < 0.001$), external regulation and reputation $H = 680.83$ (** $p < 0.001$), fairness and equality $H = 139.71$ (** $p < 0.001$), health and safety $H = 962.78$ (** $p < 0.001$), legality, constitutionality, jurisdiction $H = 68.51$ (** $p < 0.001$), morality and ethics $H = 40.89$ (** $p < 0.001$), policy prescription and evaluation $H = 1534.82$ (** $p < 0.001$), political factors and implications $H = 2776.45$ (** $p < 0.001$), public sentiment $H = 152.55$ (** $p < 0.001$), quality of life $H = 117.41$ (** $p < 0.001$), security and defense $H = 691.13$ (** $p < 0.001$).

Overall, News was the only venue that tended to contextualize LLMs in health across a wide range of social issues, including economic, policy, political, safety, and cultural considerations. On the other hand, other platforms had less comprehensive discussions, with Reddit and TikTok showing the lowest coverage of dimensions like resource, policy, or fairness. Instead, they had a significantly higher tendency to mention the impacts on quality of life than other venues, suggesting their personal and practical perspective.

4.2.2 Risk and Generative Nature Disclosure. As shown in Fig. 5, we found that public communication surrounding LLMs overall lacked a thorough introduction or overview to their risks, especially in explaining different mechanisms from traditional information sources. There was a notable asymmetry in the engagement of risk communication across venues, where layperson-driven platforms (i.e., TikTok and Reddit) were less likely to mention risks and LLM mechanisms, compared to professional and institutional sources such as News and Research Press.

The most frequently addressed risk across all platforms was related to the *information ecosystem* – how LLMs may affect people, technologies, and norms that shape the way information is produced, shared, and evaluated. A closer examination of the data showed that this type of concerns was concentrated on direct and

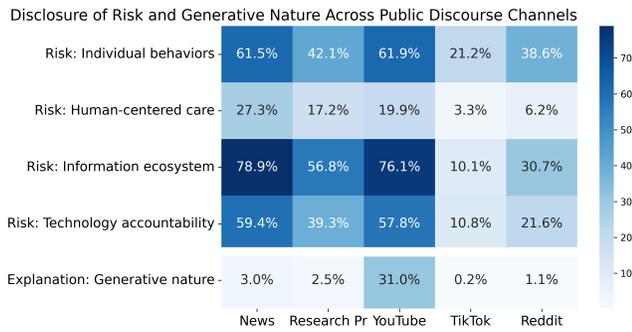


Figure 5: Disclosure of risks and generative nature in public discourse on LLMs for health. Public discourse exhibited unbalanced risk disclosure across channels, lacked thorough risk coverage, and rarely discussed the generative nature and distinctive functionalities or affordances of LLMs. Kruskal–Wallis H-tests showed significant differences across channels ($p < 0.001$, * $p < 0.01$, $p < 0.05$), where risks to individuals $H = 1564.17$ (***), risks to human-centered care $H = 1460.14$ (***), risks to information ecosystems $H = 5042.14$ (***), risks to technology accountability $H = 3213.67$ (***), and explanations of generative nature $H = 4618.17$ (***)**

isolated information quality issues. Common examples include hallucination and inaccuracies (e.g., plausible-but-wrong answers or fabricated citations), bias (e.g., racist outputs), or misuse of AI-generated content (e.g., deepfakes or AI-generated misinformation). For example, “I asked ChatGPT how to get rid of chronic pain from meniscus injuries. While its suggestions weren’t completely wrong, they weren’t exactly right either, and can worsen the situation.” This type of information quality issue was especially commonly mentioned in News (78.9%), YouTube (76.1%), and Research Press (56.8%).

The next two most commonly mentioned categories were risks to individual behaviors and risks to technology accountability, and risks to human-centered care were the least frequently mentioned. *Risks to individual behaviors* included potential negative influences on individuals’ actions, decision-making, and well-being. This type of risk was frequently mentioned in News (61.5%) and YouTube (61.9%). Common examples included impacts on employment and work norms, over-reliance on AI for decision-making, and the erosion of interpersonal or professional skills.

Similarly, *risks to technology accountability* were less prevalent on TikTok and Reddit, despite appearing with reasonable frequency in other venues. These risks refer to negative consequences related to how LLMs may challenge existing mechanisms of oversight and security. Some frequently discussed examples were algorithmic biases, unauthorized imitation of personal traits or proprietary work, and misuse for harmful purposes or human-like roles. Last, *risks to human-centered care* were least frequent. These concerns surround how LLMs may undermine the relational foundations of healthcare and disrupt existing social support systems. Despite all content touching on LLM’s potential in health, risks to the healthcare system were infrequently addressed in institutional sources and nearly absent on TikTok (3.3%) and Reddit (6.2%).

One concerning observation in our results is that explicit explanations of LLMs’ generative nature, which makes LLMs differ from conventional health information technologies, were nearly absent in all channels. Only YouTube made a moderate mention of mechanisms (31.0%), likely due to its longer-form and tutorial-based content. Across News, Research Press, Reddit, and TikTok, references to the workings of LLMs remained less than 3% or near absent (Reddit 1.1% and TikTok 0.2%), indicating a widespread lack of conceptual differentiation between LLMs and traditional search engines or decision aids – popular tools for online health information seeking in the past two decades.

4.2.3 Health Subdomain Prevalence. We identified what health subdomains in which LLMs are presented as applicable in public discussion from the inductive coding. We summarized 20 subcategories that were organized into six broader categories: clinical decision support, workflow optimization, professional training, research and data analysis, consumer health support, and healthcare system support. As shown in Fig. 6, the most commonly mentioned health subdomains were concentrated in clinical decision support (especially in diagnosis support and treatment planning and monitoring) and consumer health support (particularly in therapeutic mental health support and medical question answering). Meanwhile, the least discussed subdomains were those related to professional training and workflow optimization.

Among platforms, YouTube emerged as a middle ground between institutional sources (e.g., News and Research Press) and layperson-driven social media platforms (e.g., TikTok and Reddit). It tended to play a unique role with a large number of educational and business materials, such as talks, lectures and tutorials, which was reflected in its significantly higher content length (see Table 1) and could contribute to its broader topical coverage. This explains its focus on practical use cases, especially in healthcare data analysis, drug discovery, diagnostic and risk prediction, and medical imaging analysis. YouTube’s relatively even distribution across clinical, research, and support-oriented subdomains indicates its unique positioning that blends professional and public-facing communication.

On the other hand, the other two social media platforms, TikTok and Reddit, showed clear emphasis on consumer-facing and mental health support applications. The most frequently mentioned areas were therapeutic mental health support (TikTok: 26.5% and Reddit 21.4%) and emotional companionship (TikTok: 22.7% and Reddit 21.4%). Followed by treatment planning and monitoring, healthy lifestyle guidance, and patient communication and education. Aligned with their tendency to focus on personal and practical perspectives as observed in the framing patterns, they were less likely to discuss LLM applications in research and data analysis.

News, by comparison, emphasized clinical decision support, especially diagnostic support and risk prediction, alongside notable attention to therapeutic mental health support. Research Press followed a similar trend in highlighting clinical applications, but also had substantial coverage of research-oriented subdomains, such as drug discovery, healthcare data analysis, and novel treatment discovery, aligned with its unique role in communicating scientific and translational research findings.

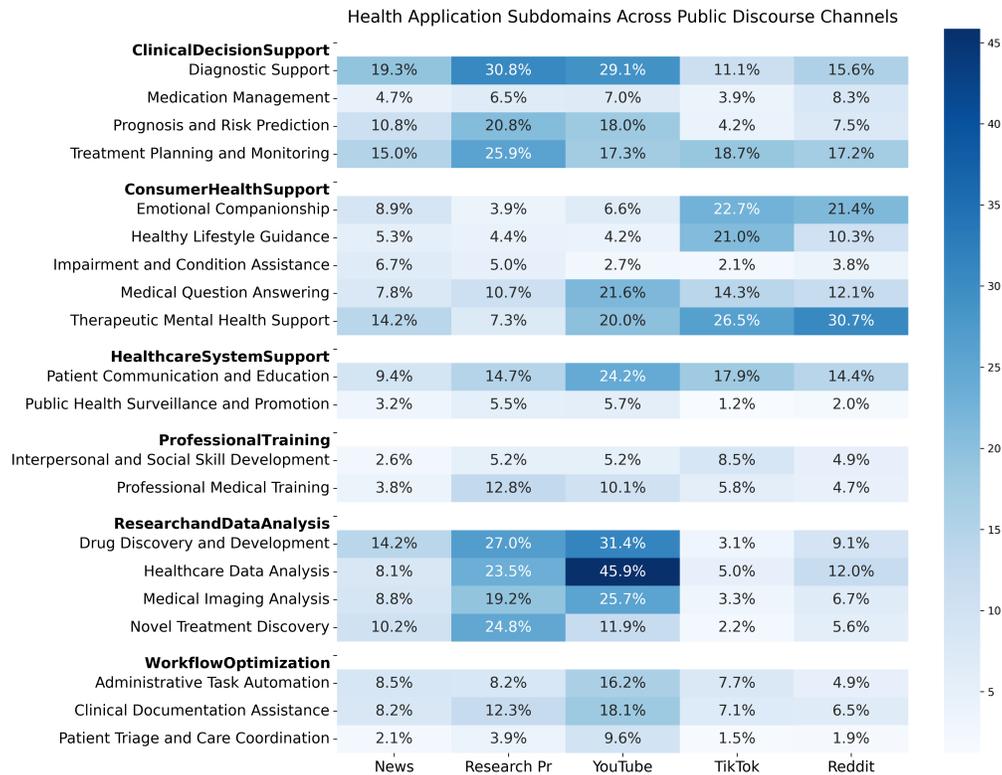


Figure 6: Health subdomains in which LLMs are mentioned in public discourse. TikTok and Reddit emphasized consumer-facing applications, particularly mental health support, while News and Research Press focused on clinical decision support and research-oriented uses. YouTube’s longer, education-oriented content contributes to broader topical coverage. Kruskal–Wallis H-tests showed significant differences only in Research and Data Analysis, where $H = 15.76 ()$ ($*** p < 0.001$, $** p < 0.01$, $* p < 0.05$).**

4.3 Symbolic Representation

Summary: Public discussions had relatively neutral to low overall anthropomorphism of LLM entities. TikTok and Reddit showed wider ranges of anthropomorphism, with some highly human-like portrayals of LLMs, while Research Press showed the significantly least anthropomorphism.

In examining the symbolic representation of LLMs, we measured the degree of anthropomorphism in describing LLM-related entities. As Fig. 7 shows, expressions across all discourse channels had interquartile ranges below zero, which means that public discussions had no apparent anthropomorphism tendencies and most direct references framed LLMs more as non-human entities. At the same time, we observed cross-platform differences in the representation of AI. Specifically, Research Press exhibited the lowest levels of anthropomorphism, which aligns with its tendency to adopt technical descriptions. In contrast, TikTok and Reddit tended to have wider ranges of anthropomorphism scores. These two platforms also contained highly anthropomorphized expressions with scores near or above 2 that described LLMs with human-like capabilities (e.g., “my ChatGPT therapist just knows what’s up in my mind”), intentions (e.g., “my chatbot cares about me”), or relationships (e.g.,

“he fell in love with and married his AI bot”). Overall, verbs associated with the highest anthropomorphism scores included “arrive”, “terrify”, “baffle”, “learn”, while common verbs mentioned in the lowest scores were “cost”, “apply”, “use”, “include”. On the other hand, News and YouTube showed an intermediate pattern of moderate and less variable anthropomorphic framing, with fewer extreme values.

5 Discussion

This paper investigates how public discourse introduces large language models, and frames their potential uses and risks in the health domain: one of the most discussed and consequential application areas. Overall, we found that discussions on LLMs for health across platforms were generally positive — and becoming more so over time, but often lacked a thorough introduction or overview of risks. When risks were mentioned, they were largely about information quality, with nearly absent explanations about the generative nature of LLMs that distinguish them from traditional information sources. Across channels, layperson-driven platforms such as TikTok and Reddit were less likely to mention risks and displayed greater variations in emotional tone and anthropomorphic expressions. They also emphasized consumer health (especially mental health support), whereas professionally authored sources, such as news and

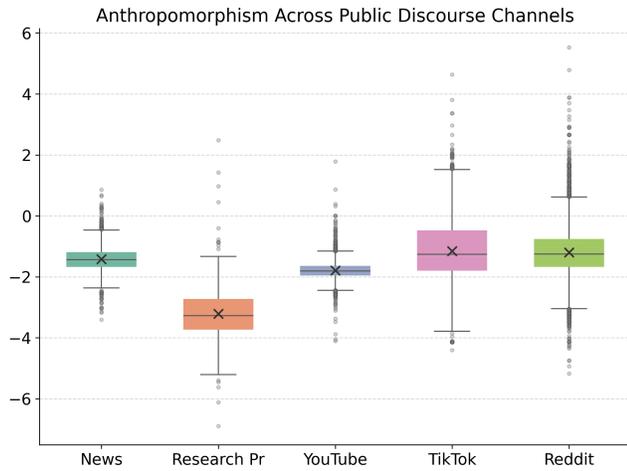


Figure 7: Anthropomorphism of LLM entities in public discourse on LLMs for health. Public discussions had relatively neutral to low overall anthropomorphism of LLM entities, with some highly human-like portrayals of LLMs in TikTok and Reddit. Kruskal–Wallis H-test found significant differences across channels, where $H = 3974.28$ (*) . Dunn’s post hoc test indicated that News was significantly less likely to adopt anthropomorphized representations of LLM entities.**

research press, put greater emphasis on clinical use and were more likely to mention broader societal or systemic implications.

Together, this work shows that discourse can work as a diagnostic tool to track public expectations and attention, as well as to identify gaps in literacy or governance. In the case of LLM-related discussions, public attitudes were found to be generally positive and increasingly so over time, but risk coverage was limited and explanations of LLMs’ generative nature were nearly absent. These findings underscore a need for effective communication and design strategies to help the public build agency and knowledge in navigating engagements with LLMs and their applications in health.

5.1 The Role of Public Discourse in Negotiating AI Perceptions

One of our primary contributions is a large-scale empirical analysis of how LLMs for health are represented and framed across diverse media channels, which we believe serves as an infrastructure for shaping public perceptions and negotiating practices. Prior research has demonstrated the role of media in communicating the risks and benefits of emerging technologies [9, 67, 109]. In terms of AI risks, scholarship has found that the media coverage, ranging from newspapers and magazines to blogs, had a shallow and mainly practical focus on AI ethics issues [89], while global news showed a skewed prioritization on societal and legal and rights-related risks [3]. Our findings extend these by documenting how different channels of public narratives reflect and reinforce this effect. Across platforms and content creators, we found that discussions were largely positive while notably lacking in mentioning risks or explaining the generative nature of LLMs. This is concerning as

LLMs’ functionalities (e.g., probabilistic generator) and affordances (e.g., general-purpose and non-task-driven design) mechanistically differ from traditional information sources.

The lack of discussion of risks and explanation of the distinctive functionalities of LLMs can encourage individuals to interact with LLMs using “old norms” that may no longer apply. Particularly at a time when current design patterns of many LLM systems often overshadow the generative nature of AI outputs or blur the boundaries between generative and retrieval systems. As technological affordances will inevitably shape users’ mental models [83], recent empirical findings showed that people held inaccurate perceptions of LLMs [118, 119]. Taken together, all evidence underscores the importance of efforts to regulate public perception and promote AI literacy. While recent regulations like the European Union (EU) AI Act [30] have emphasized the importance of AI literacy, these efforts focused on programs “intended for professionals, organisations, or the general workforce” [29]. Thus, we highlight the unique role of public discourse in engaging and influencing the greater public’s everyday perceptions and practices. After all, communicating risks should be a two-way exchange of information [54], not reduced to correcting “wrong” concerns or edifying “uninformed” lay perceptions.

A closer look at cross-channel differences reveals another important feature of the discourse infrastructure: heterogeneity among content author types. Our results showed that professional outlets (i.e., news and research press) were more likely to mention risks and broader societal and systemic implications beyond specific and isolated instances. Meanwhile, there was also an unaligned focus on specific health domains for LLMs applications in layperson outlets: TikTok and Reddit had more discussions surrounding consumer health support and especially in emotional companionship and therapeutic support. This asymmetry has implications for both information environments and information consumers. For *information environments*, it suggests that areas of high public interest (here mental health care) were under-addressed by professional outlets, leaving other channels and content that are more risk-agnostic to fill the gap. This highlights the need for journalists, researchers, and professionals to actively identify where public interest is situated, and to develop strategies for reaching audiences in those spaces with more balanced communication strategies. For *information consumers*, relying primarily on platforms like TikTok and Reddit may reduce exposure to risk-aware perspectives and increase the likelihood of developing over-optimistic or anthropomorphized views of AI. Here, a critical awareness of platform differences and tendencies becomes essential, so as to allow audiences to interpret content not only based on what is said, but also in light of the communication environments that shape what is highlighted, appreciated, or encouraged.

On a higher level, the greater variation in tone and anthropomorphism in TikTok and Reddit suggests that public perceptions of LLMs remain unsettled and actively negotiated. We see this active negotiation as both a risk and an opportunity. On one hand, selective exposure to highly positive or human-like portrayals can skew expectations and normalize inappropriate mental models. On the other hand, these unsettled framings can provide insight for journalists, designers, and policymakers to understand where public

attention is gathering, where professional communication is absent, and where new gaps in literacy or governance emerge.

5.2 Implications for AI Transparency and Governance

5.2.1 What to disclose in supporting AI transparency. In efforts to support AI transparency and literacy, the first question that follows is “what to disclose or educate”. Current regulations offer some direction: for instance, the EU AI Act requires that “developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes)” [30]. While such disclosure is important, especially as AI systems are increasingly embedded into existing systems and “disappear into” prior technologies, it may not be sufficient. Especially when categories of AI have not reached consensus in public understanding, end-users lack clarity about what capabilities and limitations different systems entail or offer. Our findings raise the question of whether informing people that they are “interacting with AI” meaningfully prepares them to interpret and evaluate their interactions.

Prior literature on risk communication underscores that effective messaging should go beyond stating the existence of risk; instead, it involves communicating not only the nature of risks but also expressing concerns, perspectives/opinions, and reactions to them [54]. Some current common practices include acknowledging error possibility [40] and uncertainty indicators [56] or reminding limitations of language models [110]. But past work in human-AI interaction has emphasized both functional understanding of what models or systems work and mechanistic understanding of how models or systems work in AI transparency [63], and showed that analytical thinking that disrupts heuristic thinking on direct statements can reduce AI overreliance [11].

However, turning to our findings, we will find a significant gap: explanations about the generative nature of LLMs were nearly missing across all channels. Even when risks were mentioned, they overwhelmingly centered on information quality, and a closer look at the data shows that the majority of the content focused on isolated incidents of fake or incorrect information. However, we question whether collapsing all information into the frequently mentioned yet broad category of hallucination is enough, when the metaphor “hallucination” itself is imperfect. LLMs do not have sensory perception as humans, so it does not naturally carry the appropriate referential analogy to explain why such errors occur. Therefore, we call attention to the gap in actively disclosing and educating on the mechanistic understanding of LLMs: how they differ from prior technologies and why these differences may lead to risks.

5.2.2 What to align in facing governance lag. In examining the specific health subdomains mentioned in public discussions, we found that these health application subdomains align broadly with categories in World Health Organization’s guidance on large multimodal models [88], but offered more nuanced typology in specific use cases. Importantly, as mentioned above, we found that compared to professional-authored content (news articles and research press releases) that emphasized clinical or systemic uses, TikTok and Reddit mentioned consumer health and mental health care more — ranging from emotional companionship to therapeutic support. This in part could be attributed to the ongoing mental health

crisis in the United States and globally, especially among youth populations [32, 87], and their pursuit of support. For instance, a survey in 2025 found that 72 percent of American teenagers said they had used AI as companions [22]. Alongside this trend, researchers have documented alarming risks, as LLMs can encourage schizophrenic delusions [60], self-harm and abusive behaviors [16, 111], and inability to handle emergency or high-stakes situations [57, 108], and even being linked to actual suicides or violent crimes [26, 111].

These documented risks, combined with our findings on high demands, point to the blurred line between informal emotional companionship and high-stakes therapeutic interventions in AI systems, which can raise safety concerns that policy frameworks have only begun to address. Some recent governance efforts in specifying boundaries of AI use for mental health care include Illinois’ Act that prohibits AI from acting as a therapist or counselor [34] and New York’s requirement for AI tools to refer people with suicidal or self-harm tendencies to crisis centers [79]. However, overall, research, practice, and policy in this area are emergent and rapidly evolving, as professional societies and stakeholders are urging AI tools to be grounded in psychological science, co-developed with behavioral health experts, and rigorously tested for safety [1].

This misalignment between public demand and regulatory clarity underscores the need for future work on both design and governance. Future work can develop strategies for in situ literacy through scaffolding or heuristic interactions to prepare users in developing mechanistic understanding, evaluating outputs, and critically trading off risks and benefits in sensitive contexts like health. On the governance side, empirical evidence from public discourse can help regulators prioritize domains where demand is high and risks are pressing or imminent. Literature in risk communication, mass media, and health education all suggest that the more effective approach is to promote better understanding and greater agency and stewardship [24, 47, 54]. We envision opportunities for future work to contribute experimental and empirical evaluations to provide better clarity as to what information most effectively supports user agency and informed engagement.

5.3 Limitations and Future Work

From a generalization perspective, this work focused on English-language content with some data (TikTok, YouTube, and News) limited to the U.S. context. While some patterns, such as the positive framing of LLMs or limited discussion of risks, can be observed in other cultural or linguistic contexts (e.g., [3]), we acknowledge that not all findings can be directly applied to non-English or non-U.S. discourse. Future work can extend this analysis to multilingual and cross-cultural settings to better understand global public perceptions of LLMs. Similarly, our categorization of risks was based on one taxonomy tailored to LLMs in the health domain. While this choice aligns with our study focus, it may not capture all potential risk dimensions relevant to other domains, and future work can utilize the methodological framework of this study to analyze public discourse on LLMs in other domains. Second, our data and analyses may be subject to platform- and source-related biases: TikTok API cutoffs may skew results toward older content, Pushshift may incompletely capture Reddit activity and introduce sampling

bias, and MBFC filtering may favor mainstream outlets. In addition, although cross-model validation demonstrated the reliable performance of gpt-4o-mini, our large-scale annotations rely on this single model and may still contain model-specific biases. Third, even with human validation, we acknowledge that human annotations of complex constructs such as health subdomains and framing remain subject to interpretive uncertainty [14] and automated labeling is inherently imperfect. Therefore, these labels should be interpreted as approximate signals rather than ground truth. Lastly, we drew on the agenda-setting theory and focused on the salience of issues and certain attributes in messages. There are additional subtle attributes that can affect audience interpretation and attitude, such as visual elements in videos and varying levels of emphasis or depth in the presentation of information elements. Mentions of LLM capabilities, limitations, or risks were treated equally in this work, regardless of whether they appeared early or late in a narrative, or were introduced superficially or in detail. We hope our study can inspire future work to examine more fine-grained aspects within discourse content and their effects on perceptions of technological capabilities and risks.

6 Conclusion

This paper examined how public discourse introduces large language models (LLMs) in the health domain, an area where concerns about reliability are emerging alongside reports linking LLM use to risky behaviors, psychological risks, and even suicide cases. Analyzing five channels (news, research press, YouTube, TikTok, and Reddit) over a 2-year period, we found that public discourse was generally positive and episodic, without risks being thoroughly communicated and often reduced to information quality incidents, and explanations of LLMs' generative nature were almost absent. Professional outlets more often addressed clinical and systemic implications, whereas layperson platforms of TikTok and Reddit highlighted mental health care and companionship, and showed greater variation in tone and anthropomorphism but little attention to risks. These findings position public discourse as a diagnostic tool for identifying where public attention gathers, where professional communication is absent, and where gaps in literacy or governance emerge. Our work contributes empirical evidence of how LLMs are currently framed in discourse and points toward future efforts to 1) support professionals in addressing gaps in information environments and information consumers in developing critical awareness of source differences; 2) develop communication strategies and in situ literacy scaffolds that move beyond surface disclosures to foster mechanistic understanding that help users critically assess capabilities and risks; and 3) guide regulatory efforts into areas with high demand but pressing risks.

Acknowledgments

Zhou and De Choudhury were partly supported through a contract from the U.S. Centers for Disease Control and Prevention (CDC). Findings reported in this paper represent the views of the authors, and not of their employers or the sponsor CDC. We thank Viet Cuong (Johnny) Nguyen, Shravika Mittal, Koustuv Saha, the Social Dynamics and Well-Being Lab members, and the anonymous reviewers for their valuable input on the paper.

References

- [1] Zara Abrams. 2025. Using generic AI chatbots for mental health support: A dangerous trend. <https://www.apaservices.org/practice/business/technology/artificial-intelligence-chatbots-therapists>
- [2] Open AI. 2022. *Introducing ChatGPT*. <https://openai.com/index/chatgpt/>
- [3] Mowafak Allaham, Kimon Kieslich, and Nicholas Diakopoulos. 2025. Informing AI Risk Assessment with News Media: Analyzing National and Political Variation in the Coverage of AI Risks. *arXiv preprint arXiv:2507.23718* (2025).
- [4] Alison Anderson, Stuart Allan, Alan Petersen, and Clare Wilkinson. 2005. The framing of nanotechnologies in the British newspaper press. *Science communication* 27, 2 (2005), 200–220.
- [5] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* 183, 6 (2023), 589–596.
- [6] Martin W Bauer. 2002. Controversial medical and agri-food biotechnology: a cultivation analysis. *Public understanding of science* 11, 2 (2002), 93.
- [7] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 830–839.
- [8] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.
- [9] Asa Boholm. 1998. Comparative studies of risk perception: a review of twenty years of research. *Journal of risk research* 1, 2 (1998), 135–163.
- [10] Amber E Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *New Directions in Analyzing Text as Data Workshop, London*. 27–28.
- [11] Zana Bućinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–21.
- [12] Felix Busch, Lena Hoffmann, Lina Xu, Long Jiang Zhang, Bin Hu, Ignacio García-Juárez, Liz N Toapanta-Yanchapaxi, Natalia Gorelik, Valérie Gorelik, Gaston A Rodriguez-Granillo, et al. 2025. Multinational attitudes toward AI in health care and diagnostics among hospital patients. *JAMA Network Open* 8, 6 (2025), e2514452–e2514452.
- [13] Paul Calle, Ruosi Shao, Yunlong Liu, Emily T Hébert, Darla Kendzor, Jordan Neil, Michael Businelle, and Chongle Pan. 2024. Towards AI-Driven Healthcare: Systematic Optimization, Linguistic Analysis, and Clinicians' Evaluation of Large Language Models for Smoking Cessation Interventions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
- [14] Dallas Card, Amber Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 438–444.
- [15] Fidelia Cascini, Ana Pantovic, Yazan A Al-Ajlouni, Giovanna Failla, Valeria Puleo, Andriy Melnyk, Alberto Lontano, and Walter Ricciardi. 2022. Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature. *EClinicalMedicine* 48 (2022).
- [16] Center for Countering Digital Hate. 2025. Fake Friend: How ChatGPT betrays vulnerable teens by encouraging dangerous behavior. <https://counterhate.com/research/fake-friend-chatgpt/>
- [17] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 1–45.
- [18] Media Bias Fact Check. [n. d.]. Media Bias/Fact Check: The Most Comprehensive Media Bias Resource. <https://mediabiasfactcheck.com/>
- [19] Myra Cheng, Kristina Gligorić, Tiziano Piccardi, and Dan Jurafsky. 2024. AnthroScore: A Computational Linguistic Measure of Anthropomorphism. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. 807–825.
- [20] Myra Cheng, Angela Y Lee, Kristina Rapuano, Kate Niederhoffer, Alex Liebscher, and Jeffrey Hancock. 2025. From tools to thieves: Measuring and understanding public perceptions of AI through crowdsourced metaphors. *arXiv preprint arXiv:2501.18045* (2025).
- [21] Bernard Cecil Cohen. 2015. Press and foreign policy. (2015).
- [22] Common Sense Media. 2025. Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions. <https://www.common SenseMedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions>
- [23] Sauvik Das, Tiffany Hyun-Jin Kim, Laura A Dabbish, and Jason I Hong. 2014. The effect of social influence on security sensitivity. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 143–157.

- [24] Sauvik Das, Joanne Lo, Laura Dabbish, and Jason I Hong. 2018. Breaking! A typology of security and privacy news and how it's shared. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [25] Jojanneke Drog, Megan Milota, Anne van den Brink, and Karin Jongmsa. 2024. Ethical guidance for reporting and evaluating claims of AI outperforming human doctors. *npj Digital Medicine* 7, 1 (2024), 271.
- [26] Maggie Harrison Dupré. 2025. Man Falls Into AI Psychosis, Kills His Mother and Himself. <https://futurism.com/man-chatgpt-psychosis-murders-mother>
- [27] William H Dutton and Grant Blank. 2011. Next generation users: the internet in Britain. (2011).
- [28] Ullrich KH Ecker, Li Qian Tay, Jon Roozenbeek, Sander Van Der Linden, John Cook, Naomi Oreskes, and Stephan Lewandowsky. 2024. Why misinformation must not be ignored. *American Psychologist* (2024).
- [29] European Parliament and Council of the European Union. 2025. AI Literacy Programs in Europe – Supporting Article 4 of the EU AI Act. <https://artificialintelligenceact.eu/ai-literacy-programs/>
- [30] European Parliament and Council of the European Union. 2024. The EU Artificial Intelligence Act. <https://artificialintelligenceact.eu/>
- [31] Jessica T Feezell. 2018. Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political research quarterly* 71, 2 (2018), 482–494.
- [32] Centers for Disease Control, Prevention, et al. 2024. Youth risk behavior survey data summary & trends report 2013 – 2023. (2024).
- [33] American Association for the Advancement of Science. [n. d.]. EurekAlert Dataset. <https://www.eurekalert.org/>
- [34] Bruce Gil. 2025. Illinois Bans AI From Providing Therapy. <https://gizmodo.com/illinois-bans-ai-from-providing-therapy-2000639042>
- [35] Mauricio Gruppi, Benjamin D Horne, and Sibel Adali. 2021. NELA-GT-2020: A large multi-labelled news dataset for the study of misinformation in news articles. *arXiv preprint arXiv:2102.04567* (2021).
- [36] Stuart Hall et al. 1997. *Representation: Cultural Representations and Signifying Practices*. Vol. 2. SAGE.
- [37] Daniel Halpern and Jennifer Gibbs. 2013. Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in human behavior* 29, 3 (2013), 1159–1168.
- [38] Stefan Harrer. 2023. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 90 (2023).
- [39] Helen Heinrichs, Alexander Kies, Saskia K Nagel, and Fabian Kiessling. 2025. Physicians' Attitudes Toward Artificial Intelligence in Medicine: Mixed Methods Survey and Interview Study. *Journal of Medical Internet Research* 27 (2025), e74187.
- [40] Kashmir Hill. 2025. They Asked an A.I. Chatbot Questions. The Answers Sent Them Spiraling. <https://www.nytimes.com/2025/06/13/technology/chatgpt-ai-chatbots-conspiracies.html>
- [41] Benjamin D Horne, Mauricio Gruppi, Kenneth Joseph, Jon Green, John P Wihbey, and Sibel Adali. 2022. NELA-Local: A dataset of US local news articles for the study of county-level news ecosystems. In *Proceedings of the international AAAI conference on web and social media*, Vol. 16. 1275–1284.
- [42] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.
- [43] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Rafford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276* (2024).
- [44] Muhammad Ittefaq, Ali Zain, Rauf Arif, Mohammad Ala-Uddin, Taufiq Ahmad, and Azhar Iqbal. 2025. Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries. *Telematics and Informatics* 96 (2025), 102223.
- [45] Shanto Iyengar. 1994. *Is anyone responsible?: How television frames political issues*. University of Chicago Press.
- [46] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15.
- [47] Nancy K Janz and Marshall H Becker. 1984. The health belief model: A decade later. *Health education quarterly* 11, 1 (1984), 1–47.
- [48] Shan Jiang and Christo Wilson. 2018. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (11 2018). doi:10.1145/3274351
- [49] Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, and Srijan Kumar. 2024. Ask Me in English Instead: Cross-Lingual Evaluation of Large Language Models for Healthcare Queries. In *The Web Conference 2024*. <https://openreview.net/forum?id=GpGSMoiBn>
- [50] Eunkyung Jo, Daniel A Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [51] Eunkyung Jo, Yun Jeong, SoHyun Park, Daniel A Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.
- [52] Kowe Kadoma, Danaë Metaxa, and Mor Naaman. 2025. Generative AI and Perceptual Harms: Who's Suspected of using LLMs?. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [53] Elise Karinschak, Sunny Xun Liu, Joon Sung Park, and Jeffrey T Hancock. 2023. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–29.
- [54] Ralph L Keeney and Detlof Von Winterfeldt. 1986. Improving risk communication. *Risk analysis* 6, 4 (1986), 417–424.
- [55] Paul M Kellstedt. 2003. *The mass media and the dynamics of American racial attitudes*. Cambridge University Press.
- [56] Sunnie SY Kim, Q Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. 2024. "I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 822–835.
- [57] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.
- [58] Gary King, Benjamin Schneer, and Ariel White. 2017. How the news media activate public expression and influence national agendas. *Science* 358, 6364 (2017), 776–780.
- [59] Heather Landi. 2024. HIMSS24: How Epic is building out AI, ambient technology for clinicians. <https://www.fiercehealthcare.com/ai-and-machine-learning/himss24-how-epic-building-out-ai-ambient-technology-clinicians>
- [60] Frank Landymore. 2025. Psychologist Says AI Is Causing Never-Before-Seen Types of Mental Disorder. <https://futurism.com/psychologist-ai-new-disorders>
- [61] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- [62] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. 2022. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110* (2022).
- [63] Q Vera Liao and Jennifer Wortman Vaughan. 2023. AI transparency in the age of llms: A human-centered research roadmap. *arXiv preprint arXiv:2306.01941* (2023), 5368–5393.
- [64] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, et al. 2023. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-radiology* 1, 2 (2023), 100017.
- [65] Travis Lloyd, Jenna Gosciak, Tung Nguyen, and Mor Naaman. 2025. AI Rules? Characterizing Reddit Community Policies Towards AI-Generated Content. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [66] Rijul Magu, Arka Dutta, Sean Kim, Ashiqur R. KhudaBukhsh, and Munmun De Choudhury. 2025. Navigating the Rabbit Hole: Emergent Biases in LLM-Generated Attack Narratives Targeting Mental Health Groups. In *Second Conference on Language Modeling*. <https://openreview.net/forum?id=am6p8VFm9l>
- [67] Leonie A Marks, Nicholas Kalaitzandonakes, Lee Wilkins, and Ludmila Zakharaova. 2007. Mass media framing of biotechnology news. *Public understanding of science* 16, 2 (2007), 183–203.
- [68] Angeline McCall. 2023. AI-powered app designed to help domestic violence victims. <https://www.9news.com/article/tech/ai-powered-app-to-help-domestic-violence-victims/73-57207485-2573-439a-b494-6da5af37d713>
- [69] Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly* 36, 2 (1972), 176–187.
- [70] Maxwell E McCombs, Donald L Shaw, and David H Weaver. 2018. New directions in agenda-setting theory and research. In *Advances in foundational mass communication theories*. Routledge, 131–152.
- [71] Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling Framing in Immigration Discourse on Social Media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2219–2263.
- [72] Microsoft. 2023. Reinventing Search with a New AI-Powered Microsoft Bing and Edge: Your Copilot for the Web. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>
- [73] Shravika Mittal, Jasmine C Foriest, Benjamin D Horne, and Munmun De Choudhury. 2024. News Media and Violence Against Women: Understanding Framings

- of Stigma. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1068–1081.
- [74] Shravika Mittal, Hayoung Jung, Mai ElSherief, Tanushree Mitra, and Munmun De Choudhury. 2025. Online myths on opioid use disorder: A comparison of reddit and large language model. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 19. 1224–1245.
- [75] Mazda Moayeri, Elham Tabassi, and Soheil Feizi. 2024. WorldBench: Quantifying Geographic Disparities in LLM Factual Recall. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1211–1228.
- [76] Vikram Mohanty, Jude Lim, and Kurt Luther. 2025. What Lies Beneath? Exploring the Impact of Underlying AI Model Updates in AI-Infused Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [77] Christian Montag, Haibo Yang, and Jon D Elhai. 2021. On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in public health* 9 (2021), 641673.
- [78] Gustavo Moreira, Edyta Paulina Bogucka, Marios Constantinides, and Daniele Quercia. 2025. The Hall of AI Fears and Hopes: Comparing the Views of AI Influencers and those of Members of the US Public Through an Interactive Platform. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–27.
- [79] New York State Governor Office. 2025. *Governor Hochul Signs Legislation Expanding Historic Empire AI Consortium and Establishing Nation-Leading AI Protections for Responsible Growth as Part of the FY 2026 Budget*. <https://www.governor.ny.gov/news/governor-hochul-signs-legislation-expanding-historic-empire-ai-consortium-and-establishing>
- [80] U.S. News. 2023. U.S. News Launches Generative AI Search Across US-News.com. <https://www.usnews.com/info/blogs/press-room/articles/2023-11-21/u-s-news-launches-generative-ai-search-across-usnews-com>
- [81] Viet Cuong Nguyen, Mini Jain, Abhijat Chauhan, Heather Jamie Soled, Santiago Alvarez Lesmes, Zihang Li, Michael L Birnbaum, Sunny X Tang, Srijan Kumar, and Munmun De Choudhury. 2025. Supporters and Skeptics: LLM-Based Analysis of Engagement with Mental Health (Mis) Information Content on Video-Sharing Platforms. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 19. 1329–1345.
- [82] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375* (2023).
- [83] Donald A. Norman. 2002. *The Design of Everyday Things*. Basic Books, Inc., USA.
- [84] Information Commissioner's Office. [n.d.]. Next-generation search. <https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/tech-horizons-report/next-generation-search/>
- [85] Jesutofunmi A Omiye, Jenna C Lester, Simon Spichak, Veronica Rotemberg, and Roxana Daneshjou. 2023. Large language models propagate race-based medicine. *NPJ Digital Medicine* 6, 1 (2023), 195.
- [86] World Health Organization. [n.d.]. Health Topics. <https://www.who.int/health-topics>
- [87] World Health Organization. 2025. Mental health of adolescents. <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>
- [88] World Health Organization et al. 2024. WHO releases AI ethics and governance guidance for large multi-modal models.
- [89] Leila Ouchchy, Allen Coin, and Veljko Dubljević. 2020. AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & society* 35, 4 (2020), 927–936.
- [90] Jeongeun Park, Changhoon Oh, and Ha Young Kim. 2024. AI vs. human-generated content and accounts on Instagram: User preferences, evaluations, and ethical considerations. *Technology in Society* 79 (2024), 102705.
- [91] Kinjal Patel, Milind Shah, Karishma M Qureshi, and Mohamed Rafik N Qureshi. 2025. A systematic review of generative AI: importance of industry and startup-centered perspectives, agentic AI, ethical considerations & challenges, and future directions. *Artificial Intelligence Review* 59, 1 (2025), 7.
- [92] James W. Pennebaker, Cindy K. Chung, Joey Frazee, Gary M. Lavergne, and David I. Beaver. 2014. When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE* 9, 12 (12 2014), e115844. doi:10.1371/JOURNAL.PONE.0115844
- [93] Snehal Prabhudesai, Ananya Prashant Kasi, Anmol Mansingh, Anindya Das Anttar, Hua Shen, and Nikola Banovic. 2025. "Here the GPT made a choice, and every choice can be biased": How Students Critically Engage with LLMs through End-User Auditing Activity. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.
- [94] Weihong Qi, Jinsheng Pan, Hanjia Lyu, and Jiebo Luo. 2024. Excitements and concerns in the post-ChatGPT era: Deciphering public perception of AI through social media analysis. *Telematics and Informatics* 92 (2024), 102158.
- [95] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.
- [96] Dejan Ravšelj, Damijana Keržič, Nina Tomažević, Lan Umek, Nejc Brezovar, Noorminshah A Iahad, Ali Abdulla Abdulla, Anait Akopyan, Magdalena Waleska Aldana Segura, Jehan AlHumaid, et al. 2025. Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLoS One* 20, 2 (2025), e0315011.
- [97] Elizabeth Reid. 2023. Supercharging Search with generative AI. <https://blog.google/products/search/generative-ai-search/>
- [98] Reuters. 2024. OpenAI says ChatGPT's weekly users have grown to 200 million. <https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/>
- [99] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, Theresa Nguyen, and Tim Althoff. 2024. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–29.
- [100] Eva Sharma, Koustuv Saha, Sindhu Kiranmai Ernala, Sucheta Ghoshal, and Munmun De Choudhury. 2017. Analyzing ideological discourse on social media: A case study of the abortion debate. In *Proceedings of the 2017 international conference of the computational social science society of the Americas*. 1–8.
- [101] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [102] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54. doi:10.1177/0261927X09351676
- [103] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* (2023), 1–11.
- [104] Tao Tu, Anil Palepu, Mike Schaeckermann, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Nenad Tomasev, et al. 2024. Towards conversational diagnostic ai. *arXiv preprint arXiv:2401.05654* (2024).
- [105] Tatjana Vasiljeva, Ilmars Kreituss, and Ilze Lulle. 2021. Artificial intelligence: the attitude of the public and representatives of various industries. *Journal of Risk and Financial Management* 14, 8 (2021), 339.
- [106] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [107] Jiajun Wang, Xiangzhe Yuan, Siying Hu, and Zhicong Lu. 2025. AI vs. Human Paintings? Deciphering Public Interactions and Perceptions Towards AI-Generated Paintings on TikTok. *International Journal of Human-Computer Interaction* (2025), 1–24.
- [108] Lu Wang, Munif Ishad Mujib, Jake Williams, George Demiris, and Jina Huh-Yoo. 2021. An evaluation of generative pre-training model-based therapy chatbot for caregivers. *arXiv preprint arXiv:2107.13115* (2021).
- [109] David A Weaver, Erica Lively, and Bruce Bimber. 2009. Searching for a frame: News media tell the story of technological progress, risk, and regulation. *Science communication* 31, 2 (2009), 139–166.
- [110] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. "As an AI language model, I cannot": Investigating LLM Denials of User Requests. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.
- [111] Joe Wilkins. 2025. Teens Keep Being Hospitalized After Talking to AI Chatbots. <https://futurism.com/teen-hospitalized-ai-chatbot>
- [112] Clare Wilkinson, Stuart Allan, Alison Anderson, and Alan Petersen. 2007. From uncertainty to risk?: Scientific and news media portrayals of nanoparticle safety. *Health, risk & society* 9, 2 (2007), 145–157.
- [113] Michael Wornow, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A Pfeffer, Jason Fries, and Nigam H Shah. 2023. The shaky foundations of large language models and foundation models for electronic health records. *npj Digital Medicine* 6, 1 (2023), 135.
- [114] Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. 2024. Hallucination is inevitable: An innate limitation of large language models. *arXiv preprint arXiv:2401.11817* (2024).
- [115] Haiyang Yang, Tinglong Dai, Nestoras Mathioudakis, Amy M Knight, Yuna Nakayasu, and Risa M Wolf. 2025. Peer perceptions of clinicians using generative AI in medical decision-making. *NPJ Digital Medicine* 8, 1 (2025), 530.
- [116] Hye Sun Yun and Timothy Bickmore. 2025. Online health information-seeking in the era of large language models: cross-sectional web-based survey study. *Journal of Medical Internet Research* 27 (2025), e68560.
- [117] Kai Zhang, Rong Zhou, Eashan Adhikarla, Zhiling Yan, Yixin Liu, Jun Yu, Zhengliang Liu, Xun Chen, Brian D Davison, Hui Ren, et al. 2024. A generalist vision-language foundation model for diverse biomedical tasks. *Nature Medicine* (2024), 1–13.
- [118] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–26.

- [119] Jiawei Zhou, Amy Z. Chen, Darshi Shah, Laura M. Schwab-Reese, and Munmun De Choudhury. 2025. A Risk Taxonomy and Reflection Tool for Large Language Model Adoption in Public Health. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article 363 (nov 2025), 32 pages. <https://doi.org/10.1145/3757544>
- [120] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3544548.3581318

A Keywords Used in Public Discourse Data Collection

LLM-relevant Keywords:

language model, llm, language artificial intelligence, language ai, generative ai, chatgpt, gpt, med-palm, google gemini, google bard, google lamda, transformer model, natural language generation, nlg, self-supervised learning, zero-shot, few-shot, generative model, autoregressive transformer, instruction tuned, meta llama, meta ai llama, meta galactica, microsoft copilot, mixtral, mistral ai, anthropic claude, deepseek, diffusion model, bloom model, qwen, xai grok

Health-relevant Keywords:

General and Symptom Terms: health, medical, disease, well being, well-being, therapy, therapist, diagnosis, diagnose, drug, nutrition, clinic, clinical, medicine, doctor, surgeon, sick, illness, disorder, patient, mental, surgical, surgery, insurance, hospital, abuse, pharma, over dose, sore, tenderness, arm pain, back pain, breast pain, ear pain, ear ache, eye irritation, flank pain, general aches, general pain, body discomfort, hand pain, headache, head pain, head discomfort, kidney pain, knee pain, leg pain, neck-skull pain, rib pain, shoulder pain, sore throat, tooth ache, fatigue, sleep disturbance, abnormal dreams, vivid dreams, asthenia, disturbed sleep, insomnia, drowsiness, lassitude, listlessness, exhaustion, lethargic, sluggishness, somnolence, tiredness, weariness, abdominal pain, abdominal cramp, anorexia, changes in appetite, bloating, constipation, difficulty swallowing, dry mouth, dyspepsia, heartburn, indigestion, epigastric pain, upset stomach, nausea, stomach ache, stomach pain, affect lability, emotional instability, irritable mood, mood changes, mood swing, depression, depressed mood, suicidal ideation, blurred vision, neuropathy, pins and needles, tingling, dizziness, lightheadedness, head rush, wooziness, angina, chest pain, difficulty breathing, dyspnea, short of breath, congestion, heart flutter, palpitation, arthralgia, joint pain, muscle pain, stiffness, amnesia, brain fog, memory loss, difficulty concentrating, short attention span, skin irritation, itchiness, pruritus, skin discomfort, dysuria, vaginal dryness, sexual dysfunction, malaise, clinician, physician, nurse, surgeon, doctor, therapist, psychology, psychological, anxiety, depression, mental health, mental illness, mental disorder, bipolar, bpd, ptsd, paranoia, schizophrenia, schizophrenic, schizo, panic attack, anxiety attack, social anxiety, self harm, self-harm, eating disorder, binge eating disorder, anorexia, anorexic, bulimia, bulimic, depressed, depressing, suicidal, suicide

Health Topics: Abortion, Addictive behaviour, Adolescent health, Anaemia, Antimicrobial resistance, Assistive technology, Biologicals, Blood products, Blood transfusion safety, Brain health, Buruli

ulcer (Mycobacterium ulcerans infection), Cancer, Cardiovascular diseases, Cervical cancer, Chagas disease, American trypanosomiasis, Chemical incidents, Chemical safety, Chikungunya, , Child health, Childhood cancer, Cholera, Chronic respiratory diseases, , Clinical trials, Common goods for health, Congenital disorders, Contraception, Coronavirus disease (COVID-19), Crimean-Congo haemorrhagic fever, Deafness and hearing loss, Dementia, Dengue and severe dengue, Depression, Diabetes, Diagnostics, Diarrhoea, Digital health, Diphtheria, Disability, Dracunculiasis (Guinea-worm disease), Injuries, Drugs, Earthquakes, Ebola virus disease, Echinococcosis, Epilepsy, Eye care, vision impairment and blindness, Female genital mutilation, Food fortification, Food safety, Foodborne diseases, Foodborne trematode infections, Health Laws, Behavioural interventions, Health promoting schools, Behavioural interventions, Health and well-being, Health promotion, Health technology assessment, Health workforce, Healthy diet, Hendra virus infection, Hepatitis, HIV, Hospitals, Human African trypanosomiasis (sleeping sickness), Human genome editing, Hypertension, In vitro diagnostics, Infant nutrition, Infection prevention and control, Infertility, Influenza (avian and zoonotic), Influenza (seasonal), Infodemic, Intellectual property and trade, International Health Regulations, Landslides, Lassa fever, Lead poisoning, Leishmaniasis, Leprosy (Hansen disease), Lymphatic filariasis (Elephantiasis), Malaria, Malnutrition, Marburg virus disease, Maternal health, Measles, Medical devices, Medicines, Meningitis, Mental health, Micronutrients, Middle East respiratory syndrome coronavirus (MERS-CoV), Mpox, Mycetoma, chromoblastomycosis and deep mycoses, Neglected tropical diseases, Newborn health, Nipah virus infection, Nursing and midwifery, Nutrition, Obesity, Occupational health, Onchocerciasis (river blindness), Oral health, Palliative care, Patient safety, Pertussis, Physical activity, Plague, Pneumonia, Poliomyelitis (polio), Primary health care, Quality of care, Rabies, Radiation, Radiation emergencies, Radon, Refugee and migrant health, Rehabilitation, Respiratory syncytial virus, Rift Valley fever, Road traffic injuries, Scabies, Schistosomiasis (Bilharzia), Self-care for health and well-being, Sepsis, Severe Acute Respiratory Syndrome (SARS), Health and wellbeing, Sexual and reproductive health and rights, Sexual health, Sexually transmitted infections (STIs), Smallpox, Snakebite envenoming, Social determinants of health, Soil-transmitted helminthiasis, Stillbirth, Substandard and falsified medical products, Suicide prevention, Sustainable development, Syphilis, Taeniasis and cysticercosis, Tetanus, Tick-borne encephalitis, Tobacco, Trachoma, Traditional, Complementary and Integrative Medicine, Transplantation, Travel and health, Tropical Cyclones, Tsunamis, Tuberculosis, Typhoid, Ultraviolet radiation, Universal health coverage, Urban health, Vaccines and immunization, Violence against children, Violence against women, Volcanic eruptions, Women's health, Yaws (Endemic treponematoses), Yellow fever, Zika virus disease

B Human–Model Validation and Cross-Model Agreement of LLM Annotations

Model	Risk Disclosure			Generative Nature			Health Subdomain		
	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
gpt-4o-mini performance	0.743	0.725	0.795	0.851	0.955	0.800	0.792	0.745	0.888
gemini-2.5 performance	0.715	0.763	0.698	0.851	0.955	0.800	0.742	0.804	0.705
cross-model agreement	0.710	0.701	0.744	0.707	0.783	0.671	0.775	0.782	0.768

Table 2: Performance of GPT-4o-mini and Gemini-2.5 against human annotations (n = 100) and inter-model agreement on a larger subset (n = 500). All scores are macro-averaged.

C Platform pairwise comparison

		Tone			Analytic		
		<i>d</i>	95% CI	<i>p</i>	<i>d</i>	95% CI	<i>p</i>
News	Research Pr	-0.46	[-0.55, -0.37]	***	-0.75	[-0.84, -0.66]	***
News	YouTube	-0.96	[-1.02, -0.91]	***	1.58	[1.52, 1.64]	***
News	TikTok	-0.66	[-0.72, -0.61]	***	1.29	[1.23, 1.35]	***
News	Reddit	-0.12	[-0.17, -0.08]	***	0.6	[0.55, 0.64]	***
Research Pr	News	0.46	[0.37, 0.55]	***	0.75	[0.66, 0.84]	***
Research Pr	YouTube	-0.48	[-0.56, -0.39]	***	1.74	[1.65, 1.83]	***
Research Pr	TikTok	-0.26	[-0.34, -0.17]	***	1.45	[1.36, 1.54]	***
Research Pr	Reddit	0.22	[0.14, 0.30]	***	0.85	[0.77, 0.93]	***
YouTube	News	0.96	[0.91, 1.02]	***	-1.58	[-1.64, -1.52]	***
YouTube	Research Pr	0.48	[0.39, 0.56]	***	-1.74	[-1.83, -1.65]	***
YouTube	TikTok	0.06	[0.01, 0.11]	***	-0.16	[-0.21, -0.12]	***
YouTube	Reddit	0.54	[0.50, 0.58]	***	-0.63	[-0.67, -0.60]	***
TikTok	News	0.66	[0.61, 0.72]	***	-1.29	[-1.35, -1.23]	***
TikTok	Research Pr	0.26	[0.17, 0.34]	***	-1.45	[-1.54, -1.36]	***
TikTok	YouTube	-0.06	[-0.11, -0.01]	***	0.16	[0.12, 0.21]	***
TikTok	Reddit	0.46	[0.42, 0.50]	***	-0.47	[-0.51, -0.43]	***
Reddit	News	0.12	[0.08, 0.17]	***	-0.6	[-0.64, -0.55]	***
Reddit	Research Pr	-0.22	[-0.30, -0.14]	***	-0.85	[-0.93, -0.77]	***
Reddit	YouTube	-0.54	[-0.58, -0.50]	***	0.63	[0.60, 0.67]	***
Reddit	TikTok	-0.46	[-0.50, -0.42]	***	0.47	[0.43, 0.51]	***

		Clout			Authentic		
		<i>d</i>	95% CI	<i>p</i>	<i>d</i>	95% CI	<i>p</i>
News	Research Pr	0.39	[0.30, 0.48]	***	0.28	[0.19, 0.37]	***
News	YouTube	-0.49	[-0.55, -0.44]	***	-0.62	[-0.67, -0.57]	***
News	TikTok	-0.36	[-0.42, -0.31]	***	-0.46	[-0.51, -0.40]	***
News	Reddit	0.52	[0.47, 0.56]	***	-0.49	[-0.53, -0.44]	***
Research Pr	News	-0.39	[-0.48, -0.30]	***	-0.28	[-0.37, -0.19]	***
Research Pr	YouTube	-0.85	[-0.94, -0.76]	***	-0.84	[-0.93, -0.76]	***
Research Pr	TikTok	-0.53	[-0.62, -0.44]	***	-0.56	[-0.65, -0.47]	***
Research Pr	Reddit	0.33	[0.25, 0.41]	***	-0.58	[-0.66, -0.50]	***
YouTube	News	0.49	[0.44, 0.55]	***	0.62	[0.57, 0.67]	***
YouTube	Research Pr	0.85	[0.76, 0.94]	***	0.84	[0.76, 0.93]	***
YouTube	TikTok	-0.04	[-0.09, 0.00]	***	-0.03	[-0.08, 0.02]	***
YouTube	Reddit	0.78	[0.74, 0.82]	***	-0.2	[-0.23, -0.16]	***
TikTok	News	0.36	[0.31, 0.42]	***	0.46	[0.40, 0.51]	***
TikTok	Research Pr	0.53	[0.44, 0.62]	***	0.56	[0.47, 0.65]	***
TikTok	YouTube	0.04	[-0.00, 0.09]	***	0.03	[-0.02, 0.08]	***
TikTok	Reddit	0.76	[0.72, 0.80]	*	-0.17	[-0.20, -0.13]	***
Reddit	News	-0.52	[-0.56, -0.47]	***	0.49	[0.44, 0.53]	***
Reddit	Research Pr	-0.33	[-0.41, -0.25]	***	0.58	[0.50, 0.66]	***
Reddit	YouTube	-0.78	[-0.82, -0.74]	***	0.2	[0.16, 0.23]	***
Reddit	TikTok	-0.76	[-0.80, -0.72]	*	0.17	[0.13, 0.20]	***

Table 3: Platform pairwise comparison for lexical styles. Cohen’s *d* with Dunn’s test *p*-values (Bonferroni corrected) where * $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Across platforms, YouTube showed a significantly more positive tone, Research Press showed significantly higher analyticity, TikTok showed significantly greater clout, and Reddit showed significantly higher authenticity.**

		Risks to individuals			Risks to human-centered care			Risks to information ecosystems		
		<i>d</i>	95% CI	<i>p</i>	<i>d</i>	95% CI	<i>p</i>	<i>d</i>	95% CI	<i>p</i>
News	Research Pr	0.4	[0.31, 0.49]	***	0.23	[0.15, 0.32]	***	0.52	[0.43, 0.61]	***
News	YouTube	-0.01	[-0.06, 0.04]	***	0.18	[0.12, 0.23]	***	0.07	[0.01, 0.12]	***
News	TikTok	0.91	[0.86, 0.97]	***	0.76	[0.70, 0.82]	***	1.98	[1.91, 2.04]	***
News	Reddit	0.47	[0.43, 0.52]	***	0.75	[0.70, 0.79]	***	1.06	[1.02, 1.11]	***
Research Pr	News	-0.4	[-0.49, -0.31]	***	-0.23	[-0.32, -0.15]	***	-0.52	[-0.61, -0.43]	***
Research Pr	YouTube	-0.41	[-0.49, -0.32]	***	-0.07	[-0.15, 0.02]	***	-0.44	[-0.53, -0.36]	***
Research Pr	TikTok	0.49	[0.41, 0.58]	***	0.62	[0.54, 0.71]	***	1.37	[1.28, 1.46]	***
Research Pr	Reddit	0.07	[-0.01, 0.15]	***	0.44	[0.36, 0.52]	***	0.56	[0.48, 0.64]	***
YouTube	News	0.01	[-0.04, 0.06]	***	-0.18	[-0.23, -0.12]	***	-0.07	[-0.12, -0.01]	***
YouTube	Research Pr	0.41	[0.32, 0.49]	***	0.07	[-0.02, 0.15]	***	0.44	[0.36, 0.53]	***
YouTube	TikTok	0.9	[0.85, 0.95]	***	0.53	[0.48, 0.58]	***	1.78	[1.72, 1.84]	***
YouTube	Reddit	0.48	[0.44, 0.52]	***	0.48	[0.44, 0.52]	***	1	[0.96, 1.04]	***
TikTok	News	-0.91	[-0.97, -0.86]	***	-0.76	[-0.82, -0.70]	***	-1.98	[-2.04, -1.91]	***
TikTok	Research Pr	-0.49	[-0.58, -0.41]	***	-0.62	[-0.71, -0.54]	***	-1.37	[-1.46, -1.28]	***
TikTok	YouTube	-0.9	[-0.95, -0.85]	***	-0.53	[-0.58, -0.48]	***	-1.78	[-1.84, -1.72]	***
TikTok	Reddit	-0.37	[-0.41, -0.33]	***	-0.13	[-0.16, -0.09]	***	-0.48	[-0.52, -0.44]	***
Reddit	News	-0.47	[-0.52, -0.43]	***	-0.75	[-0.79, -0.70]	***	-1.06	[-1.11, -1.02]	***
Reddit	Research Pr	-0.07	[-0.15, 0.01]	***	-0.44	[-0.52, -0.36]	***	-0.56	[-0.64, -0.48]	***
Reddit	YouTube	-0.48	[-0.52, -0.44]	***	-0.48	[-0.52, -0.44]	***	-1	[-1.04, -0.96]	***
Reddit	TikTok	0.37	[0.33, 0.41]	***	0.13	[0.09, 0.16]	***	0.48	[0.44, 0.52]	***

		Risks to technology accountability			Explanations of generative nature		
		<i>d</i>	95% CI	<i>p</i>	<i>d</i>	95% CI	<i>p</i>
News	Research Pr	0.41	[0.32, 0.50]	***	0.03	[-0.06, 0.11]	**
News	YouTube	0.03	[-0.02, 0.09]	***	-0.74	[-0.80, -0.69]	***
News	TikTok	1.23	[1.17, 1.29]	***	0.25	[0.19, 0.30]	***
News	Reddit	0.89	[0.84, 0.94]	***	0.16	[0.11, 0.20]	***
Research Pr	News	-0.41	[-0.50, -0.32]	***	-0.03	[-0.11, 0.06]	**
Research Pr	YouTube	-0.37	[-0.46, -0.29]	***	-0.66	[-0.75, -0.58]	***
Research Pr	TikTok	0.82	[0.74, 0.91]	***	0.31	[0.23, 0.40]	***
Research Pr	Reddit	0.43	[0.35, 0.51]	***	0.13	[0.05, 0.21]	***
YouTube	News	-0.03	[-0.09, 0.02]	***	0.74	[0.69, 0.80]	***
YouTube	Research Pr	0.37	[0.29, 0.46]	***	0.66	[0.58, 0.75]	***
YouTube	TikTok	1.13	[1.08, 1.18]	***	0.92	[0.87, 0.97]	***
YouTube	Reddit	0.84	[0.80, 0.88]	***	1.25	[1.21, 1.29]	***
TikTok	News	-1.23	[-1.29, -1.17]	***	-0.25	[-0.30, -0.19]	***
TikTok	Research Pr	-0.82	[-0.91, -0.74]	***	-0.31	[-0.40, -0.23]	***
TikTok	YouTube	-1.13	[-1.18, -1.08]	***	-0.92	[-0.97, -0.87]	***
TikTok	Reddit	-0.28	[-0.31, -0.24]	***	-0.1	[-0.14, -0.06]	***
Reddit	News	-0.89	[-0.94, -0.84]	***	-0.16	[-0.20, -0.11]	***
Reddit	Research Pr	-0.43	[-0.51, -0.35]	***	-0.13	[-0.21, -0.05]	***
Reddit	YouTube	-0.84	[-0.88, -0.80]	***	-1.25	[-1.29, -1.21]	***
Reddit	TikTok	0.28	[0.24, 0.31]	***	0.1	[0.06, 0.14]	***

Table 4: Platform pairwise comparison for informational content: disclosure of risks and generative nature. Cohen's *d* with Dunn's test *p*-values (Bonferroni corrected) where * $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.**

		Anthropomorphism		
		<i>d</i>	95% CI	<i>p</i>
News	Research Pr	3.21	[3.09, 3.33]	***
News	YouTube	1	[0.94, 1.05]	***
News	TikTok	-0.28	[-0.33, -0.22]	***
News	Reddit	-0.28	[-0.32, -0.23]	***
Research Pr	News	-3.21	[-3.33, -3.09]	***
Research Pr	YouTube	-3.18	[-3.29, -3.07]	***
Research Pr	TikTok	-1.85	[-1.95, -1.75]	***
Research Pr	Reddit	-2.43	[-2.52, -2.35]	***
YouTube	News	-1	[-1.05, -0.94]	***
YouTube	Research Pr	3.18	[3.07, 3.29]	***
YouTube	TikTok	-0.77	[-0.82, -0.72]	***
YouTube	Reddit	-0.8	[-0.84, -0.76]	***
TikTok	News	0.28	[0.22, 0.33]	***
TikTok	Research Pr	1.85	[1.75, 1.95]	***
TikTok	YouTube	0.77	[0.72, 0.82]	***
TikTok	Reddit	0.04	[0.01, 0.08]	***
Reddit	News	0.28	[0.23, 0.32]	***
Reddit	Research Pr	2.43	[2.35, 2.52]	***
Reddit	YouTube	0.8	[0.76, 0.84]	***
Reddit	TikTok	-0.04	[-0.08, -0.01]	***

Table 5: Platform pairwise comparison for symbolic representation. Cohen's *d* with Dunn's test *p*-values (Bonferroni corrected) where * $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.**

D Temporal Trends

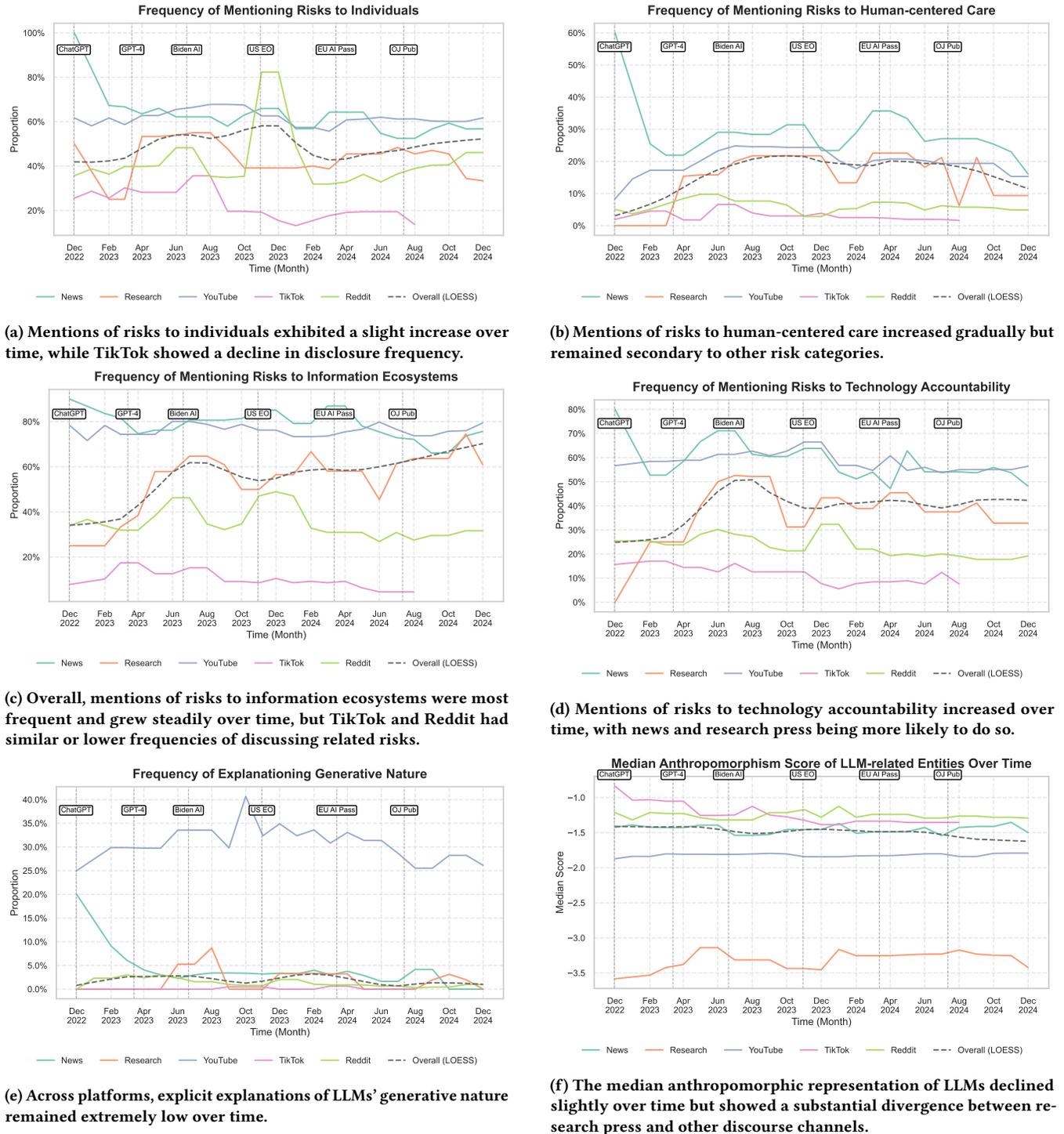


Figure 8: Temporal trends. TikTok data ended in August 2024 due to an unresolved API internal error (see Sec. 3.1.1). We applied LOESS (Locally Estimated Scatterplot Smoothing) to capture overall trends across five data sources of varying size and nature.