



Auditing TikTok's recommendations: Understanding the prevalence of misinformation on the platform

Suwani Gunasekara^{a,*}, Saumya Pareek^a, Ryan M. Kelly^b, Vassilis Kostakos^a, Jorge Goncalves^a

^a School of Computing and Information Systems, The University of Melbourne, Parkville, Melbourne, 3010, Victoria, Australia

^b School of Computing Technologies, RMIT University, Melbourne, 3000, Victoria, Australia

HIGHLIGHTS

- Employed a sock puppet audit to assess misinformation prevalence across TikTok's search & FYP algorithms.
- New accounts with no prior history are more susceptible to misinformation recommendations.
- Engaging with misinformation on one topic acts as a gateway to more misinformation on related and unrelated topics.
- Searching neutral terms can expose users to misinformation, as misleading suggestions often rank higher than debunking ones.

ARTICLE INFO

Keywords:

Misinformation
Algorithmic audits
Short video platforms
TikTok
Videos
Recommendations

ABSTRACT

Short video platforms have rapidly increased in popularity, bringing forth a concerning trend: widespread dissemination of misinformation partially owing to the persuasive nature of the rich media these platforms support. However, little research has systematically investigated the role of algorithmic recommendations in this phenomenon. Our study addresses this gap by auditing the recommendation algorithm of TikTok, the leading short video platform, through four experiments. We explore the prevalence of misinformation in TikTok's auto-complete suggestions, search results, and For You Page (FYP) recommendations. Our audits yielded a novel dataset comprising 8535 auto-complete suggestions, 1606 search queries, and 2893 videos labelled by stance (*promoting* misinformation, *neutral* in nature, or *debunking* misinformation). Our analysis reveals that TikTok's algorithm exhibits sensitivity to user behaviours, often in unexpected ways. We find the algorithm recommends substantially more misinformation to newly created accounts lacking search histories, potentially shaping the tone of the content these accounts engage with. We also find that watching and engaging with videos (e.g., liking and favouriting) dynamically influence misinformation prevalent on the platform, potentially reinforcing existing beliefs and contributing to echo-chambers. Further, we emphasise how engaging with misinformation on one topic serves as a gateway, prompting the algorithm to recommend misinformation spanning both related and unrelated topics. Our work stands as the first large-scale, systematic, and empirical assessment of misinformation in TikTok's recommendations, contributing valuable insights on the dynamics of misinformation dissemination on TikTok and also underscoring the pressing need for responsible auditing practices across digital platforms.

1. Introduction

Social media platforms have been inundated with misinformation, which often includes rich media, such as images and videos, in addition to text (Flintham et al., 2018; Spezzano et al., 2021). A glaring example is the 2020 U.S. election, where nearly one in four political image posts on Facebook contained inaccurate information (Lytvynenko, 2020).

Images and videos often trigger what psychologists term the *realism heuristic*, encapsulated by the saying “seeing is believing” (Sundar et al., 2021). Previous work has shown that content presented in audio-visual formats tends to be more favourably perceived than its textual counterpart (Appiah, 2006; Sundar, 2000). In essence, the *richer* the modality, the more likely the message will be perceived by end-users as genuine

* Corresponding author.

Email address: gunasekara@student.unimelb.edu.au (S. Gunasekara).

<https://doi.org/10.1016/j.ijhcs.2026.103929>

Received 7 October 2025; Received in revised form 4 May 2026; Accepted 18 August 2026

Available online 3 September 2026

1071-5819/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and credible. This heightened sense of realism frequently leads users to bypass their usual cognitive filters for assessing the trustworthiness and accuracy of both a piece of content and its source. Notably, videos are characterised by their sensory appeal, vividness, and direct engagement, which make them considerably more convincing in comparison to other modalities that hinge on interpretation and abstraction (Yadav et al., 2011). Thus, content modality can play a pivotal role in the spreading and curbing of misinformation on social media platforms.

Understanding the influence of content modality is particularly important given that individuals increasingly rely on social media platforms as their primary source of news. For instance, half of the adults in the U.S. obtain their news at least occasionally from social media (Liedke and Wang, 2023). In recent years, TikTok had the most pronounced increase among all social media platforms in the percentage of users who regularly obtain news from the platform, rising from 22% in 2020 to 43% in 2023. As of January 2024, TikTok boasts over 1.5 billion active users (Iqbal, 2023), mostly due to the ease of content production and the compelling nature of short-form videos.

Despite claims by various media outlets accusing TikTok of spreading misinformation (Paul, 2022; Wong, 2022; Lundy, 2023), research on misinformation within short video applications has been limited to understanding the video content posted on these platforms across different topics, such as medicine (O'Sullivan et al., 2022; Madraki et al., 2021) or politics (Boulianne and Lee, 2022; Rogers, 2021), and also the characteristics of these videos (Madraki et al., 2021; van Kampen et al., 2022). However, the role of TikTok's recommendation algorithm, a pivotal factor in content distribution (Weimann and Masri, 2020), remains understudied in relation to the prevalence of misinformation on the platform. Importantly, this algorithm does not operate independently of user activity on the platform (Zhao, 2021). As the algorithm is designed to incorporate user activities, several questions arise, such as: What types of user activities cause the recommendation algorithm to recommend misinformation to users? Does this pattern change based on the user's past activity on the platform?

Therefore, our work consists of a set of experiments to audit TikTok's recommendation algorithm: two SEARCH Experiments, exploring the algorithm's behaviour when a user searches for specific types of content, and two ENGAGEMENT Experiments, investigating the influence of user interactions on video recommendations. These experiments were conducted on three popular scientific misinformation topics: *Flat Earth*, *Moon Landing*, and *Vaccine Controversies*. Scientific misinformation is defined as publicly available information which is misleading and deceptive. Such statements contradict entities that adhere to scientific principles and are not supported by the best available scientific evidence (Southwell et al., 2022). Nevertheless, despite scientific research providing evidence against these theories, certain groups of individuals continue to passionately believe in and perpetuate them. Furthermore, these topics exhibit unique characteristics. For example, despite its age, the topic *Flat Earth* periodically resurfaces with new theories. In contrast, *Moon Landing* conspiracy theories have maintained a relatively consistent narrative since the late 1960s. *Vaccine Controversies*, on the other hand, are relatively new, and the consequences of misinformation surrounding them are severe compared to the other two topics, as it is more directly connected to public health and individuals' daily lives. We intentionally selected topics with varying characteristics to examine whether the platform's behaviour remains consistent across different types of misinformation, rather than being driven by the properties of a single topic. By examining the platform through the lens of these three topics, we contribute insights into how TikTok handles both sensitive contemporary issues and longstanding, established topics. Analysing multiple topics ensures that the findings are not unduly influenced by the characteristics of a single topic, thereby enhancing the robustness and generalizability of the results. Through our experiments, we address the following research questions:

RQ1 [SEARCH Experiments]: How does TikTok's recommendation algorithm influence the degree of misinformation presented to users during searches?

RQ1a: When a search query contains a term associated with a misinformation topic, what is the prevalence of misinformation within the auto-complete suggestions?

RQ1b: When a search query contains a term associated with a misinformation topic, what is the prevalence of misinformation within the search results?

RQ2 [ENGAGEMENT Experiments]: How do user engagement actions on TikTok influence the degree of misinformation presented to users?

RQ2a: How does a user's watch time for misinformation videos influence the content subsequently displayed on their For You Page?

RQ2b: How does user engagement with videos, such as liking and favouriting, influence the content displayed on the For You Page?

To answer the above research questions, we employed a sock-puppet audit methodology, which involves using bots to simulate user behaviour in a predefined controlled environment. This method has been widely applied to audit algorithms of online platforms for various purposes, such as measuring political biases (Le et al., 2019), price discrimination (Mikians et al., 2012; Hannak et al., 2014), and gender biases (Eriksson and Johansson, 2017; Asplund et al., 2020) in recommendation systems. For our experiments, we scraped recommendations from three sources: auto-complete suggestions, search results, and video content on TikTok's For You Page (FYP). These recommendations were then labelled based on their relevant stance (e.g., as *promoting* misinformation, *neutral* in nature, or as *debunking* misinformation). Finally, we created four account types, including new accounts and accounts with varying search histories to examine the recommendations that TikTok algorithmically provides. These histories were established by systematically searching for content using queries with the stances stated above.

Our findings show that new accounts (as opposed to accounts with a search history) tend to receive relatively more misinformation than *debunking* recommendations in the form of auto-complete suggestions and search results. Moreover, when merely seeking information on a controversial topic, there is a higher prevalence of algorithmically recommended misinformation in auto-complete suggestions and search results, across all account types and topics. Not only did the proportion of misinformation *promoting* auto-complete suggestions and search results surpass those of *debunking* recommendations across all topics and account types, but this problematic content was also positioned more prominently on the platform.

Furthermore, we show that once TikTok accounts engage with videos *promoting* misinformation, the algorithm injects more recommendations that contain misinformation on the For You Page, specifically within the first 25 videos. However, the prevalence of misinformation varies across topics. Specifically, accounts encountered a lower volume of misinformation on *Vaccine Controversies* compared to the other two topics being investigated. Interestingly, watching misinformation videos related to *Vaccine Controversies* led the algorithm to recommend videos that promote misinformation spanning other unrelated topics, serving as a gateway for more fabricated content being displayed to accounts.

This paper presents several key contributions. Firstly, we introduce a methodological pipeline for auditing short video platforms, offering a framework to quantify the effect of user actions on the prevalence of misinformation on these platforms. By applying our methodology to TikTok, our study is the first to systematically explore the influence of user actions on misinformation prevalent in auto-complete suggestions, search results, and video recommendations on a short video platform. Second, we provide a novel dataset of 8535 auto-complete suggestions, 1606

search queries, and 2893 videos labelled with their relevant stances (*promoting, debunking, neutral and irrelevant*), related to the topics: *Flat Earth, Moon Landing and Vaccine Controversies*. This dataset stands as a valuable resource for the research community to further investigate how TikTok recommends misinformation to users. Third, we provide valuable insights into the mechanisms that contribute to the prevalence and positioning of misinformation on short-video platforms such as TikTok, contributing to a broader understanding of the inner workings of content recommendation algorithms in the age of rich media consumption. Finally, we provide an in-depth discussion of our findings, underscoring the necessity for public scrutiny of recommendation algorithms to ensure platform accountability. Such scrutiny is vital for upholding transparency, ensuring ethical practices, and safeguarding user interests in an increasingly data-driven digital landscape.

2. Related work

2.1. Misinformation on video sharing platforms (VSPs)

Misinformation on social media platforms has become rampant through multimedia posts shared among users (Lytvynenko, 2020; Jane Lytvynenko, 2020). These modalities, particularly videos, offer heightened levels of *richness* in terms of available cues (Tseng et al., 2022). In comparison to text, video is considered a richer medium due to its provision of audiovisual information (Qi et al., 2022). In the context of misinformation, individuals tend to perceive video-based misinformation as more credible than textual or audio formats, which also leads to a heightened likelihood of sharing such content (Sundar et al., 2021).

Importantly, consumption of information through video has become increasingly popular due to the emergence of online video platforms such as YouTube and TikTok (Bu et al., 2023; Cuesta-Valiño et al., 2022). These platforms have a large audience, meaning that any information disseminated can quickly spread to the masses, potentially influencing public opinion and behaviour. As an example, researchers discovered that many individuals started to believe the Earth is flat after watching popular recommended videos on YouTube (Alex Olshansky, 2018). Furthermore, a recent analysis of 124 TikTok videos on misinformation related to COVID-19 vaccines revealed that these videos had generated more than 20 million views (O'Connor, 2021). Despite the existing body of literature suggesting that video format surpasses text in its effectiveness for spreading misinformation (Ecker et al., 2022; Micallef et al., 2022), research dedicated to investigating the prevalence and spread of misinformation on VSPs remains scarce.

Hussein et al. (2020) investigated how personalisation factors affect the amount of misinformation content recommended by YouTube. They identified that demographics and the geo-location do not play a significant role in amplifying misinformation in search results until the user has accumulated a watch history. Building upon this, Tomlein et al. (2021) found that a more balanced recommendation system could be achieved by actively engaging with content that debunks misinformation. Previous work by Papadamou et al. (2020) further showed that a user's watch history significantly impacts subsequent recommendations made by the YouTube algorithm. Moreover, Spinelli and Crovella (2020) established that YouTube's recommendation algorithm tends to steer users away from trustworthy information sources. This effect is amplified when users choose to keep their personal information hidden from the platform.

Importantly, misinformation studies conducted on VSPs have primarily focused on YouTube, while less attention has been directed towards short video platforms like TikTok. These platforms are particularly susceptible to the dissemination of misinformation due to the ease of content production and consumption (Bakan, 2023; O'Connor, 2021; Hsu, 2022; Tucker, 2022).

2.1.1. Short video applications

Short-form videos, typically spanning from mere seconds to a concise 5-minute duration, have redefined the video content landscape (Yang

et al., 2019). These brief yet impactful videos, predominantly created by laypeople, offer straightforward and engaging content that resonates with diverse audiences. This transformative trend challenges conventional long-form videos and has proven highly effective in achieving communication goals. TikTok stands as a prime example of such platforms, setting itself apart from established social media giants such as YouTube due to its unique short-form video model.

YouTube, with its long-form videos, also provides users with a greater control over their viewing experience by offering a range of recommendations to choose from. On the contrary, TikTok streamlines the experience by presenting recommendations without user selection. Capitalising on short-form video content, TikTok has garnered exceptional user engagement compared to its competitors. The emergence of services like Instagram Reels,¹ YouTube Shorts,² and Netflix's Fast Laughs,³ which emulate this bite-sized video style, further underscores TikTok's influential impact on the online digital landscape.

However, despite its popularity, TikTok has faced allegations of misinformation propagation (Hsu, 2022; Tucker, 2022; Paul, 2022). This holds significant importance, especially given recent studies indicating that the younger demographic is increasingly relying on TikTok as their main platform for consuming news (Tomasik and Matsa, 2025). The platform's limited user control over content recommendations and its relatively lower barrier for posting materials have sparked concerns, necessitating further investigation. TikTok's success hinges on its recommendation algorithm, yet empirical analyses of this algorithm remain notably scarce (Boeker and Urman, 2022). Importantly, among major social media platforms, TikTok remains the least investigated platform concerning the dissemination of misinformation (Yeung et al., 2022), underscoring the need for comprehensive research in light of its widespread usage and potential for misinformation spread. Therefore, in this work, we aim to contribute to the understanding of TikTok's recommendation algorithm and its implications for misinformation dissemination, based on a variety of potential user actions within the platform.

2.2. Algorithmic auditing

The underlying algorithms of recommendation systems are one of the main factors that ultimately determine what specific content users end up interacting with. Therefore, these algorithms constitute one of the most crucial components of the platforms (Brown et al., 2021), possessing the power to shape users' thoughts and opinions (Juneja, 2023). For instance, a study found that Facebook's News Feed algorithm can influence user emotions by selectively showing positive or negative posts (Xie et al., 2022). Given the substantial impact of recommendation algorithms on user perceptions, researchers have emphasised the crucial need for auditing online platforms to ensure alignment with legal, societal, ethical, and industry standards (Juneja et al., 2023).

Audits encompass the collection of data concerning an algorithm's behaviour within a specific context. Subsequently, the data is used to evaluate whether the algorithm's behaviour has adverse effects on the interests or rights of individuals influenced by the algorithm (Bandy, 2021). Researchers have suggested and applied a wide array of audit methodologies, encompassing techniques like code audits, scraping audits, sock-puppet audits, and crowdsourced audits (Bandy, 2021). A *code audit* entails analysing the algorithm's code and design for problematic behaviour. However, it is often infeasible for online platforms due to their proprietary code and the need for human experts to decipher complex logic (Pasquale, 2010, 2011). *Scraping audits* involve collecting data directly from web pages or APIs, but may overlook user-specific behaviours that influence algorithmic outputs (Sandvig et al., 2014). *Sock-puppet audits* employ non-human bots that mimic real users to study algorithmic behaviour, offering control over manipulation and

¹ <https://about.instagram.com/features/reels>.

² https://www.youtube.com/intl/en_id/creators/shorts.

³ <https://about.netflix.com/en/news/fast-laughs-is-here>.

data collection (Sandvig et al., 2014). Lastly, *crowdsourced audits* utilise crowdworkers to gather data, but may similarly introduce false or harmful data, and also come with high recruitment costs (Sandvig et al., 2014).

Previous works have employed these audit techniques to delve into various concerns relevant to algorithmically operated online platforms. For instance, studies have audited algorithmic systems to understand the presence of partisan bias in search engine platforms (Hu et al., 2019; Metaxa et al., 2019), biases concerning race and gender in on-line freelance marketplaces (Hannak et al., 2014), the lack of diversity of news sources in search engines (Trielli and Diakopoulos, 2019), and the effect of personalisation factors on web search (Hannak et al., 2013; Kliman-Silver et al., 2015). However, auditing VSPs for algorithmically propagated misinformation is relatively under-explored, apart from the few studies conducted on YouTube focusing on its long-format videos (Hussein et al., 2020; Tomlein et al., 2021; Papadamou et al., 2020; Spinelli and Crovella, 2020). Given the increasing popularity of the short-video format, it has become equally crucial to understand how algorithms disseminate misinformation on platforms predominantly offering this type of content, such as TikTok.

In this context, our study adopts the *sock-puppet auditing* technique, utilising non-human bots to emulate user behaviour in a predefined manner. This method provides us with greater flexibility in managing user interactions, a level of control that is challenging to achieve through *crowdsourcing* or *scraping audits*. *Crowdsourcing audits* present challenges in maintaining a comparable environment across users, while *scraping audits* limit the extent of simulating diverse user behaviours (Srba et al., 2023). The *sock-puppet auditing* technique addresses these limitations, allowing for a more controlled and nuanced examination of user interactions on the platform.

The work most closely related to ours is the empirical exploration conducted by Boeker and Urman (2022). The authors simulated user actions such as following and liking, and examined the influence of TikTok's recommendations within users' For You Page. They found that following users had more impact on what the user sees, compared to liking content. Additionally, they observed that both language and location have an effect on videos recommended in the FYP. Our study distinguishes itself from Boeker and Urman (2022)'s work in several significant ways. While both studies aim to understand TikTok's recommendation algorithm, ours takes a more specialised approach. Specifically, we employ a strategy involving distinct account types with varying levels of engagement with different types of misinformation related content. By doing so, we aim to provide a more thorough understanding of the algorithm's behaviour, within the specific context of misinformation, a factor not investigated in Boeker and Urman (2022)'s study. Our research also extends Boeker and Urman (2022)'s work by examining multiple sources of recommendations (auto-complete suggestions, search results and recommendations on the For You Page) on TikTok. This broader scope allows us to gain insights into algorithmic changes that may occur on various parts of the platform, a facet not specifically addressed in their study.

3. Method

Our objective in this study was to understand the prevalence of misinformation on TikTok by auditing its recommendation algorithm. Specifically, we aimed to understand how users might be exposed to misinformation through auto-completed query suggestions and search results (RQ1), and to examine how users' subsequent interactions with TikTok videos impact the prevalence of misinformation the platform presents to them (RQ2). To achieve this, we identified suitable search queries to serve as seed queries for generating auto-complete suggestions. Additionally, we created multiple TikTok accounts, with search histories reflecting specific content *stances*. In the following sections, we describe our methodology for account creation, bot implementation, and

establishing the search histories for these accounts. Subsequently, we detail our two SEARCH experiments and two ENGAGEMENT experiments. We then outline the procedures and heuristics employed for labelling queries and videos throughout the audit experiments.

3.1. Experiments overview

When a user enters TikTok, they can search for content by first typing the desired term in the search bar. This leads to the user encountering auto-complete suggestions, and they can decide to either click on them or execute the query to obtain the search results. Additionally, they can explore recommendations presented by the platform via the For You Page. Once the user selects a specific video to watch, they may engage with the content by watching it for a desired duration and potentially interacting with the video, such as liking, favouriting, and/or performing both of these actions together. To assess how these user actions could potentially contribute to the dissemination of misinformation and answer our research questions, we conducted four experiments. Before initiating these experiments, we first had to create TikTok accounts and their histories.

3.1.1. Account creation and classification

The TikTok accounts were created using Google email addresses. Hussein et al. (2020) found evidence supporting the notion that demographic attributes exert an effect on amplifying misinformation on YouTube. Therefore, to mitigate similar undue effects across our TikTok accounts, we ensured that all accounts shared identical demographic attributes. To mitigate any potential age-related biases, all email accounts shared the same birth year, ensuring a consistent age of 30 years across them. At the time of this research, the age group with the maximum number of daily TikTok users for the previous year (2022) was 25–34. To ensure representation, we selected the median age (30) within this group (Enberg, 2022). Gender information was deliberately set to “rather not say” and we used gender neutral names to prevent any biases based on gender. To control for the potential influence of following behaviour on content recommendations, as highlighted in prior work (Boeker and Urman, 2022), all accounts in our study did not follow any users nor did they have any followers throughout the duration of the experiments. This ensured that differences in recommendations were not confounded by variations in following activity.

We created four distinct account types to investigate how the recommendation algorithm's behaviour is influenced by the *stance* of the content that the account has searched for in the past. In this work, we define three content stances that are of interest to us: whether a video promotes misinformation, debunks misinformation, or is neutral in nature and does not exhibit a clear bias (see Fig. 3b). Consequently, the four account types we created were **NEW** (no prior search history), **PROMOTING [P]** (sought videos that promote misinformation), **DEBUNKING [D]** (sought videos that debunk misinformation), and **NEUTRAL [N]** (sought neutral videos on controversial topics, without a clear bias towards promoting or debunking misinformation). Further, the latter three account types with prior search histories were created for each misinformation topic examined in this study, namely, *Moon Landing*, *Flat Earth*, and *Vaccine Controversies*.

In total, 108 accounts were created, covering three topics, four account types (NEW, PROMOTING, DEBUNKING and NEUTRAL), and nine scenarios (S1–S9) to facilitate the audit experiments. Each **scenario** represents a distinct user action, such as searching queries (S2) and watching videos for different durations (S3–S6), and was designed to align with the objectives of different RQs, as outlined in Table 1. We note that typing queries (S1) and searching queries (S2) are distinct actions. Scenario S1 involves typing search queries into TikTok's search bar, allowing us to examine potential auto-complete suggestions, while S2 entails executing search queries on TikTok, enabling us to collect the subsequent search results returned by TikTok. Upon completion

Table 1

List of scenarios examined in our four audit experiments, once for each topic. NEW represents new accounts with no prior search history, while P, D, and N represent accounts with a search history focused on seeking videos that promote misinformation, debunk it, and are neutral in nature without a clear bias, respectively. Overall, nine scenarios were explored, each representing a distinct user action pivotal to a specific RQ. The final column presents the TikTok data examined to answer these RQs.

Experiment	Account Type	Scenarios	Data Examined
1. Search (RQ1a)	NEW, P, D, N	Typing queries (S1)	Auto-complete suggestions
2. Search (RQ1b)	NEW, P, D, N	Searching queries (S2)	Search results
3. Engagement (RQ2a)	NEW, P, D, N	Watching: 25% (S3), 50% (S4), 75% (S5), 100% (S6)	Recommendations on the FYP
4. Engagement (RQ2b)	NEW, P, D, N	Actions: Like (S7), Favourite (S8), Like & Favourite (S9)	Recommendations on the FYP

USER	Generate five example search queries related to the topic "Moon Landing", for each of the following distinct stances: "promoting" (seeking information supporting/confirming moon landing conspiracy theories), "debunking" (seeking factual information disproving/countering moon landing conspiracy theories), and "neutral" (general queries without a clear bias to seek information that promotes or debunks moon landing conspiracy theories). Keep the queries short.
ASSISTANT	Promoting Moon Landing conspiracy theories: 1. "Moon Landing fake evidence" 2. "NASA Moon Landing fake proof" 3. "Proof Moon Landing was staged" 4. "Moon Landing staged by Hollywood" 5. "Moon Landing faked by NASA" Debunking Moon Landing conspiracy theories: 1. "Scientific proof of Moon Landing" 2. "Proof that Moon Landing was real" 3. "Debunking Moon Landing conspiracy theories" 4. "Astronauts refute Moon Landing hoax" 5. "Moon Landing hoax debunked" Neutral queries related to Moon Landing: 1. "Moon Landing date" 2. "Apollo missions timeline" 3. "Astronauts who walked on the moon" 4. "Moon Landing mission objectives" 5. "Impact of Moon Landing on space exploration"

Fig. 1. An example prompt to GPT-4 to generate search queries for the topic *Moon Landing*. We subsequently utilised these queries to establish search histories with specific stances for our TikTok accounts.

of the experiment, all TikTok accounts created for the study were deleted.

3.1.2. Automating accounts

We utilised a Selenium bot coded in Python to manage each TikTok account. To ensure a clean and isolated browsing environment for them, and to eliminate any noise arising from tracked cookies or browser history, these bots accessed TikTok within incognito mode. The bots logged into each TikTok account using the account's credentials, carried out the scripted scenarios, and exported specific results and datasets relevant to our research questions. Each bot was located in the same geographic area to mitigate any biases arising from their geo-location. Furthermore, the data scraping process was executed in parallel, using machines with identical specifications. This approach was implemented to reduce any temporal effects resulting from variations in machine speed, ensuring consistency in the data collection process.

3.1.3. Establishing account search histories

To create a PROMOTING (P), DEBUNKING (D), or NEUTRAL (N) account, we needed to establish prior search histories aligned with these stances. This process was repeated from scratch for each of the three

misinformation topics. We utilised OpenAI's GPT-4⁴ to generate search terms typically used to seek content with promoting, debunking, and neutral stances on each of the three misinformation topics (Fig. 1). The generated search terms were then utilised to establish individualised search histories for each account type. For each account, search terms of a stance corresponding to the account's type were searched on TikTok using our bots. After searching each term, TikTok's search results page was scrolled for 30 s, mimicking a more human-like browsing behaviour compared to no scrolling at all. The bot then proceeded to search another query.

3.2. SEARCH experiments (RQ1)

3.2.1. Experiment 1: auto-complete suggestions (RQ1a)

A critical set of algorithmic recommendations that users encounter while searching queries on TikTok is the auto-complete suggestions. These suggestions are designed to predict and display possible completion options for the user's search query as it is being typed, aiming to reduce typing effort. While it may facilitate quick searches, users may

⁴ <https://openai.com/blog/chatgpt>.

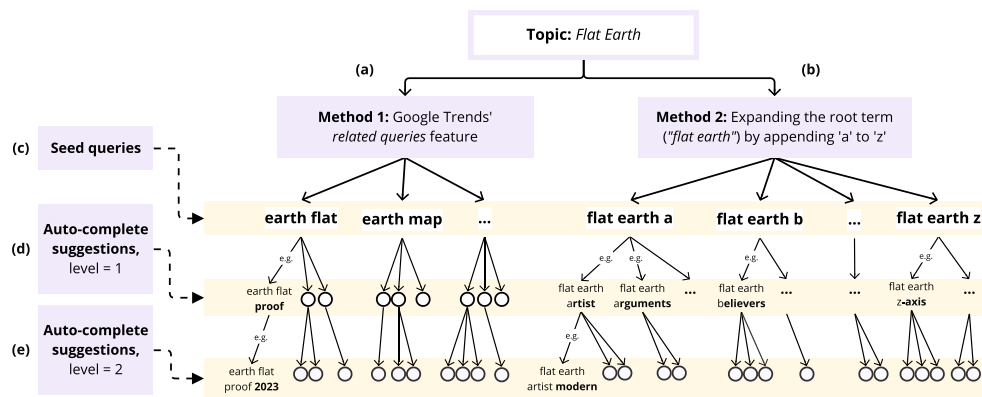


Fig. 2. The two methods utilised to generate auto-complete suggestions on TikTok from our seed queries, illustrated for the *Flat Earth* topic as an example. The first method utilised Google Trends' *related queries* feature to collect seed queries, while the second appended alphabets 'a' to 'z' to the root term ('flat earth' in this example) to iteratively generate more seed queries. These seed queries were then searched on TikTok to collect auto-complete suggestions (*level = 1*), which were then re-entered into the search box to generate a subsequent set of auto-complete suggestions (*level = 2*).

click on these suggestions even if they depart from what users intended to initially search before viewing the algorithm's auto-complete suggestions (Houli et al., 2021). Consequently, users can be influenced by the auto-complete suggestions presented to them, leading to a shift in their initial search intentions (Roy and Ayalon, 2019), altering the content subsequently consumed by them. Furthermore, the auto-complete feature has the potential to not only reinforce existing beliefs but also propagate these beliefs to new individuals, as users often prioritise popularity over accuracy on social media platforms (Allem, 2020). Therefore, in this experiment, we investigate how search queries seeking content with different stances (PROMOTING, DEBUNKING, and NEUTRAL) related to three misinformation topics (*Moon Landing*, *Flat Earth*, and *Vaccine Controversies*) display potentially "problematic" auto-complete suggestions to TikTok accounts. For this study, we define a problematic query as one which likely promotes misinformation and directs users towards such types of non-factual content.

Implementation: We used seed queries from which to generate auto-complete suggestions on TikTok. We employed two distinct methods to generate initial seed queries, as illustrated in Fig. 2. The first method involved utilising Google Trends⁵ *related queries* feature, a valuable source for identifying popular queries searched by users in a specific geographical location during a given time period. For each misinformation topic, *Moon Landing*, *Flat Earth*, and *Vaccine Controversies*, we utilised the *related queries* feature by searching these terms as a *topic* instead of a *term*. Searching as a *term* returns results containing either one or more of the search strings, while *topic* treats the entire search term as one string (Hussein et al., 2020). We set categories to 'All' and search service to 'Web Search'. Subsequently, we extracted all related queries in the U.S., spanning from January 1, 2021, to December 31, 2022. The topics' interest-over-time plot for this date range signalled that they had been persistently discussed in the last two years, which was beneficial for our research objectives. We viewed related topics using the 'Top' filter which presents the most popular search queries, and collected the top 25 related queries for each misinformation topic that was being audited (see Fig. 2a), resulting in a total set of 75 queries.

In the second method, we took a distinct approach by systematically expanding each topic's root term (see Fig. 2b). We define "root term" as the base or primary keyword of each topic, such as 'flat earth', 'moon landing', and 'vaccine', from which variations or extensions can be subsequently generated. For instance, starting with the root term 'flat earth', the expansion process involved appending alphabets starting from 'a' to 'z' to generate terms like 'flat earth a', 'flat earth b', and so on. This

systematic generation of variations of the root term aimed to create a broad spectrum of seed queries. As an illustrative example, the expanded query 'flat earth a' might lead to specific suggestions from the platform such as 'flat earth arguments' and 'flat earth artists,' demonstrating the sequential expansion of our seed query pool, as depicted in Fig. 2.

After undertaking both processes, we obtained a diverse seed query pool (see Fig. 2c) from which to generate auto-complete suggestions. To generate these suggestions, we employed recursive algorithmic interrogation techniques, enabling us to construct a suggestion tree that iteratively expanded our auto-complete suggestions and resulted in a more diverse and comprehensive dataset. The procedure involved inputting our seed queries into the TikTok search box and collecting the resulting auto-complete suggestions (*level = 1*). This process was repeated recursively for up to two levels of depth. The initial set of auto-complete suggestions obtained (*level = 1*, Fig. 2d) was then inputted into the search box again, and the auto-complete suggestions stemming from these new queries (*level = 2*, Fig. 2e) were then recorded. This approach aimed to simulate real-life search query patterns where users may utilise auto-complete suggestions to further refine their search, allowing us to accumulate a diverse set of auto-complete suggestions. This process of collecting auto-complete suggestions was applied to all seed queries for each misinformation topic represented in scenario S1 in Table 1, and is illustrated in Fig. 3(c).

3.2.2. Experiment 2: query search results (RQ1b)

In addition to auto-complete suggestions, another form of algorithmic recommendations encountered by TikTok users during searches is the results of the conducted search. In general, search engines are designed to deliver 'relevant' content (Akbar et al., 2023) and can significantly shape users' opinions (Kulshrestha et al., 2017). However, several studies have pointed out that these search results may be biased in different ways (Abualsaud and Smucker, 2019; Akbar et al., 2023; Hu et al., 2019). In RQ1b, we aim to understand how queries searched on TikTok may result in problematic search outcomes.

Implementation: For each misinformation topic, each of the four account types (NEW, PROMOTING, DEBUNKING, NEUTRAL) searched 15 distinct queries on TikTok, representing scenario S2 in Table 1. The process of selecting search queries involved a systematic procedure. After categorising the auto-complete suggestions from the first experiment (Section 3.2.1) based on their stance (promoting, neutral, and debunking), the five queries with the highest frequency in each stance category were chosen. This resulted in a set of 15 final queries, which were then shuffled and executed in sequence (see Fig. 3d). Following each search, the bot entered a sleep period of 20 min to mitigate potential carryover effects, aligning with recommendations from prior algorithmic auditing

⁵ <https://trends.google.com/trends/>.

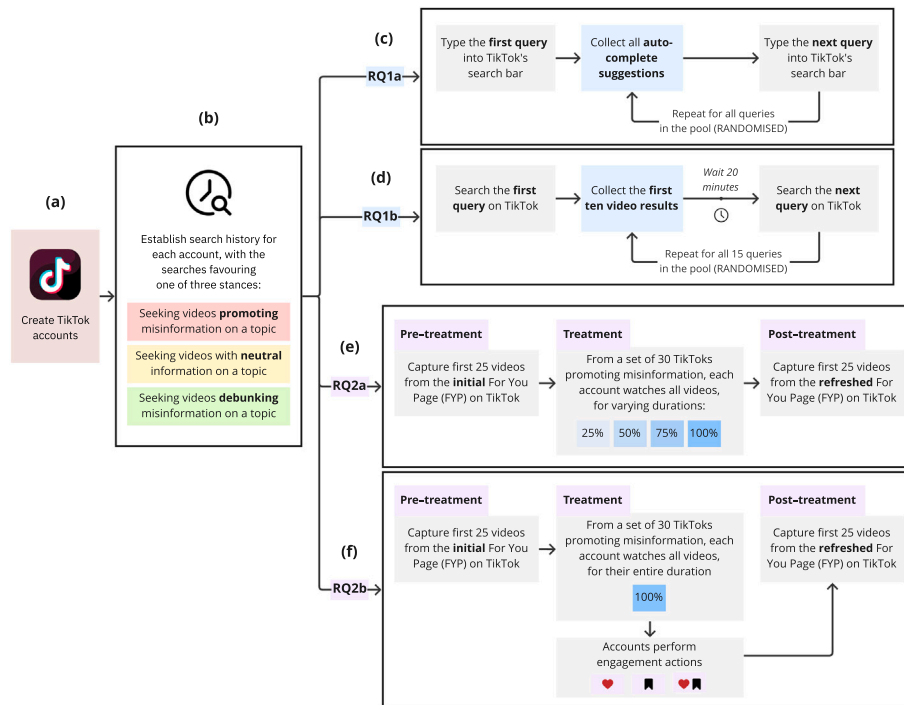


Fig. 3. Overview of the different steps performed in each of the four experiments.

experiments (Hussein et al., 2020; Srba et al., 2023). The choice of 15 queries for each topic was driven by practical considerations, as it is not feasible to search through all scraped suggestions while maintaining a 20-minute interval between each search. We collected a total of 1800 search results (15 search queries per topic $\times$ 10 videos per query $\times$ 4 account types per topic $\times$ 3 topics).

3.3. Engagement experiments (RQ2)

3.3.1. Experiment 3: video watch time (RQ2a)

The duration for which a user watches a video is generally considered a valuable indicator of user interest and engagement (Qiu and Cui, 2011; Dobrian et al., 2011). The longer a user watches a particular video, the higher the likelihood that they find the content engaging and interesting. Notably, YouTube has successfully incorporated this metric into its recommendation algorithm, utilising the percentage of a user's video watch time to recommend related videos, which has led to a 50% overall increase in watch time (Newton, 2017). Recognising the importance of this feature on YouTube, it is crucial to investigate whether a similar factor plays a pivotal role in TikTok's recommendation algorithm and whether it contributes to the dissemination of misinformation on the platform.

Implementation: As illustrated in Fig. 3(e), after logging into the accounts created for this experiment, relevant account histories were established. The experimental methodology involved three phases, illustrated in Fig. 3(e). In the pre-treatment phase, the URLs of the first 25 videos on the accounts' For You Pages (FYPs) were extracted. Subsequently, during the treatment phase, accounts watched 30 videos with a promoting stance (videos promoting misinformation and conspiracy theories related to our three misinformation topics). These videos were watched up to specified percentages of their total length (25%, 50%, 75%, and 100%), corresponding to distinct scenarios (S3–S6 respectively) in Table 1. Post-treatment, the URLs of the top 25 videos were once again extracted after refreshing the FYP. We ensured that the 30 promoting videos selected for each topic differed from those used to create the promoting watch histories, and that the selected videos did not

exceed a duration of 120 s. This procedure was executed separately for each misinformation topic. For each of the three misinformation topics, the experiment was repeated four times (once for each watch percentage: 25%, 50%, 75%, 100%) for each of the four account types, resulting in a total of 48 runs.

3.3.2. Experiment 4: engagement indicators: like, favourite, and like & favourite (RQ2b)

In addition to watch time, liking and favouriting videos serve as proxies to gauge user engagement with video content. However, despite being considered indicators of user interest and engagement, it remains unclear how short video applications respond when these indicators suggest that the user is interested in misinformation-related content. In this fourth and final experiment, we set out to answer this question. This experiments' accounts carry out three types of actions: only liking the video (*Like*), only favouriting the video (*Favourite*), and both liking and favouriting the video (*Like & Favourite*). These actions represent distinct scenarios (S7–S9 respectively) in Table 1. Through the analysis of TikTok's landing page, i.e., the For You Page (FYP), both before and after these engagement actions were executed, we sought to gain a deeper understanding of how these user actions influence the algorithm's recommendations of misinformative content.

Implementation: Similar to Experiment 3 addressing RQ2a, Experiment 4 also comprised three phases, illustrated in Fig. 3(f). During the pre-treatment phase, we extracted the URLs of the first 25 videos on the FYP by following a methodology similar to Experiment 3. Subsequently, during the treatment phase, accounts watched 30 videos with promoting stances, utilising the same set of videos from Experiment 3. These videos were fully watched, i.e., with a watch percentage of 100%. Next, accounts executed specific engagement actions on each video, namely, (*Like*), (*Favourite*), and (*Like & Favourite*). Separate TikTok accounts were created for each action type, so for example, (*Like*) and (*Favourite*) actions were undertaken by different accounts. Post-treatment, the URLs of the top 25 videos were once again extracted after refreshing the FYP.

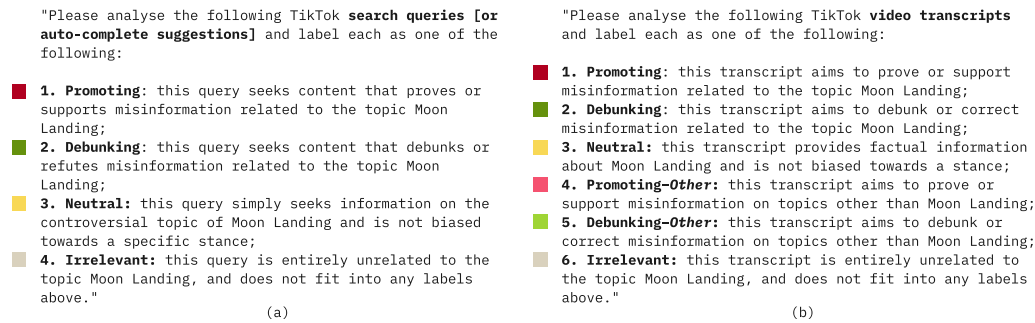


Fig. 4. ChatGPT prompt used to label (a) queries and auto-complete suggestions; and (b) videos related to the topic *Moon Landing*.

3.4. Ethical considerations

The objective of this study is to understand the prevalence of misinformation on TikTok. By quantifying misinformation and examining how these platforms propagate it to different user types, we can develop effective strategies to combat it, and inform policy and education efforts. We acknowledge the potential ethical concerns associated with utilising bots on the platform as part of our research. However, we argue that the benefits of understanding the prevalence and dynamics of misinformation far outweigh these potential risks. Importantly, we have meticulously designed the study and made specific decisions to ensure that any engagement with misinformation is conducted responsibly and with minimal risk to users. For instance, due to the nature of dealing with misinformation materials, we utilised bots instead of real users to conduct the research. Using human participants in a controlled setting would have resulted in their repeated exposure to misinformation, which we aimed to avoid. Furthermore, the number of interactions the bots had with the platform was relatively low. As a result, our actions did not cause any disruptions to the platform and are unlikely to have influenced the propensity for it to deliver misinformation to users. To understand what type of content the bot encountered, the content which was visible to each bot was scraped from TikTok. This process complied with TikTok's Terms of Service, which explicitly state that content scraping is not prohibited unless done for commercial purposes (TikTok, 2021). By prioritising the understanding of misinformation's prevalence, we aim to contribute valuable knowledge that can ultimately foster a safer and more informed digital environment for all users.

3.5. Labelling queries, auto-complete suggestions, and videos

We collected 8535 auto-complete suggestions, 1606 search queries, and 2893 unique videos. Our experimental objectives required all these to be labelled based on their *stance*, i.e., whether they sought to promote misinformation, debunk it, or were neutral in nature. Given the substantial volume of data requiring labelling, we leveraged Large Language Models (LLMs) for this task. Large Language Models (LLMs) have garnered significant interest for their potential to automate a wide range of annotation tasks (Aubin Le Quéré et al., 2024; Choi and Ferrara, 2023; Park et al., 2024; Wang et al., 2024; Heseltine and Clemm von Hohenberg, 2024; Rathje et al., 2024).

Generative LLMs such as OpenAI's GPT-3 and GPT-4 have demonstrated considerable efficacy in text labelling tasks across various Natural Language Processing (NLP) applications, outperforming several other LLMs available in the market (Caramancion, 2023). In certain instances, these models have even surpassed the performance of traditional labelling methods (Ding et al., 2022). For example, ChatGPT surpassed trained human labelers in identifying the political affiliations of Twitter users (Gilardi et al., 2023; Törnberg, 2023). Additionally, ChatGPT has exhibited the ability to classify news as either true or false with high accuracy (Hoes et al., 2023a,b). Moreover, it has demonstrated proficiency in providing accurate information regarding common cancer

myths and misconceptions (Johnson et al., 2023). These findings underscore the potential utility of LLMs in efficiently and accurately labelling content (Alizadeh et al., 2023). However, the overall performance of LLMs does not achieve 100% accuracy. It is important to note that expecting perfect agreement may be unrealistic, especially given that agreement rates among human fact-checkers are themselves not flawless (Lim, 2018). Furthermore, the decision to use the GPT-4 version was made since it is more capable of following instructions due to its ability to search online, unlike GPT-3.5 (OpenAI, 2023).

We utilised OpenAI's API *GPT-4* using zero-shot classification, instructing the model to label data based on instructions we provided. Zero-shot learning enables models to make predictions for tasks they have not been explicitly trained on, enhancing their adaptability to unseen challenges (Alizadeh et al., 2023). Furthermore, OpenAI models are non-deterministic, implying that identical inputs can result in different outputs.⁶ To regulate this, a parameter known as the *temperature* is configured, which governs the level of randomness in a model's output. This value ranges from 0 to 2, with higher values introducing more randomness and lower values yielding more focused and deterministic outputs. In this work, the *temperature* parameter was set to zero for more deterministic results. While setting the *temperature* parameter to 0 increases determinism in outputs, a slight degree of variability may still persist (Alizadeh et al., 2023).

In the following sections, we detail the procedure followed to label our collection of queries, auto-complete suggestions, and videos gathered from TikTok. Our dataset is available to the research community.⁷

3.5.1. Labelling queries and auto-complete suggestions

To formulate an LLM prompt which is representative of our collected data, we needed to gain a comprehensive understanding of our queries. We manually reviewed 100 randomly sampled queries from each misinformation topic, and labelled them based on their stance. These labels were then cross-verified using information from reputable websites and journals such as CNN, BBC, World Health Organization and PubMed Central (Dewan, 2022; Morrow, 2022; bbc, 2012; Organisation, 2020; Gabis et al., 2022). To ensure accuracy, we engaged in discussions to reach a consensus on the final label value for each text in the sample dataset. Using this curated sample dataset, we formulated and iterated upon an instruction prompt (Fig. 4a), and used the *GPT-4* API (hereinafter referred to as 'the model') to also label the randomly selected query dataset for each topic, independent of the authors' labelling. Table 2 displays the inter-rater reliability (IRR) between the model and researcher generated labels for each topic, calculated using Cohen's Kappa. The scores range from 0.66 to 0.77. These IRR scores reflect a high level of agreement and reliability (McHugh, 2012), affirming the validity and effectiveness of our refined prompt.

⁶ <https://platform.openai.com/docs/api-reference/chat/create>.

⁷ https://osf.io/z7k5u/?view_only=5623ab5563fc44ebbc03e2c64eadec19.

Table 2

Inter-Rater-Reliability (IRR) scores between the model and researcher-generated labels for each misinformation topic.

Data Labelled	Topic	IRR score
Queries and Auto-complete Suggestions	Flat Earth	0.67
	Moon Landing	0.77
	Vaccine Controversy	0.66
Search Result Videos	Flat Earth	0.68
	Moon Landing	0.66
	Vaccine Controversy	0.86

Labelling Criteria: To facilitate the model’s labelling process, we outlined specific rules in our prompt for **four labels**—*promoting*, *debunking*, *neutral*, and *irrelevant*. Queries and auto-complete suggestions were labelled as **promoting** if they sought content that supports misinformation or conspiracy theories related to the investigated topics. *Promoting* auto-complete suggestions included examples such as “*earth is flat*”, “*vaccines cause baby autism*”, and “*moon landing fake flag*”. Conversely, we instructed the model to label queries as **debunking** if they sought content that debunks or refutes misinformation related to our topic. For example, suggestions such as “*first moon landing is real*” and “*vaccine is safe*” suggest factually accurate information to users, and attempt to debunk any conspiracy theories related to the topic. Next, the model was instructed to label queries as **neutral** if they simply sought information on a controversial topic and were not biased towards a specific stance. *Neutral* suggestions included those directing users to discussions or debates related to the misinformation topic, such as “*Apollo 11 footage*” and “*flat earthers versus scientists*”. These suggestions aimed to provide factual information or facilitate discussions without endorsing or debunking conspiracy theories. Finally, queries and auto-complete suggestions which were unrelated to the topic under investigation and did not fit any of the aforementioned categories, such as “*first man camila cabello*” and “*flat minecraft seed world for building*”, were labelled as **irrelevant**.

3.5.2. Labelling videos

We asked the model to analyse the transcripts of the collected videos. We generated transcripts using OpenAI’s Whisper package (Ai, 2022), which is a state-of-the-art automatic speech recognition (ASR) system. Whisper is trained on a large dataset of multilingual and multitask supervised data, and it incorporates advanced machine learning techniques to accurately convert spoken words into written text, enabling us to effectively transcribe video content. We familiarised ourselves with the videos, and manually labelled a randomly selected set of 20 videos for each topic. We then refined the prompt through an iterative process similar to that used for queries (see Fig. 4b for the prompt). Table 2 displays the inter-rater reliability (IRR) between the model and researcher-generated labels for videos of each topic, with scores ranging between 0.66 and 0.86. Once again, the IRR scores indicate strong agreement and reliability (McHugh, 2012), confirming the validity and efficacy of our refined prompt.

Labelling Criteria: For videos, we specified **six labels**—*promoting*, *debunking*, *neutral*, *irrelevant*, *promoting-other*, and *debunking-other*. The model was instructed to label videos as **promoting** if they aimed to prove or support misinformation or conspiracy theories related to the topic under investigation. For instance, a video titled “*Flights prove the earth is flat!*”, advocating the Flat Earth theory, appeared among the top 10 videos displayed by TikTok when the search query “flat earth” was executed. Such videos were labelled as **promoting**. Further, videos received a **debunking** label if they presented information which debunked and corrected misinformation or conspiracy theories, for example, by presenting scientific evidence related to the topic under investigation. For instance, a video titled “*Debunking Moon Landing myths*” which explained reasons behind the U.S. flag waving on the moon was labelled as **debunking**. Next, the model assigned a **neutral** label to videos which

presented information objectively, without endorsing or promoting a specific stance, and were not skewed in nature. These videos often focused on discussions, debates, or providing general information, for example, “*Earth facts*”.

Additionally, upon manual inspection of randomly selected videos collected for the User-Engagement experiments (Experiments 3 and 4), we identified certain instances where videos promoted or debunked misinformation that was unrelated to the topic being investigated. For instance, accounts engaging with the topic *Flat Earth* as part of the experimental procedure received recommendations in their For You Page (FYP) promoting or debunking misinformation on the topic *Moon Landing*. To address such scenarios during video labelling, we included instructions in our prompt to label such videos as **promoting-other** and **debunking-other**, based on whether they promoted or debunked unrelated misinformation. Finally, videos which did not fit into any of the above categories were labelled as **irrelevant**.

Around 5% of the videos could not be transcribed by Whisper due to the lack of audio. Therefore, we manually labelled these videos to obtain the complete final dataset.

3.6. Analysis

We conducted an analysis of the labelled audit data to investigate how frequently and prominently misinformation is algorithmically promoted and disseminated on TikTok. Specifically, we audited auto-complete suggestions (RQ1a), search results after executing queries (RQ1b), and recommendations populated on the For You Page (RQ2a and RQ2b). For each experiment, we performed a **proportion analysis** where we calculated proportion of content suggested or returned by the platform that promotes misinformation. We compared the results of this proportion analysis across the 3 topics (*Flat Earth*, *Moon Landing* and *Vaccine Controversies*) and 4 account types (NEW, PROMOTING, DEBUNKING, NEUTRAL). We followed this with a **prominence analysis** where we examined the salience and visibility of misinformation, i.e., its positioning within auto-complete suggestions (RQ1a), search results (RQ1b), and recommendations on the For You Page (RQ2a and RQ2b). This analysis provides insights into whether misinformation was more prominent at the top, bottom, or elsewhere in the three lists—the auto-complete suggestions, search results, and FYP recommendations. The prominence score was calculated by evaluating where the misinformation appeared within these lists; the more prominent the positioning, the higher the score. In the prominence score equation, x is a boolean value indicating whether that particular position in the list contains a video on a specific stance, r indicates the position (rank) of the video within the list, and n represents the total number of items in the list. t represents a topic from *Flat Earth*, *Moon Landing* and *Vaccine Controversies*, and s represents a stance from promoting, debunking, and neutral.

$$\text{ProminenceScore}_{(topic=t, stance=s)} = \frac{\sum_{r=0}^{n-1} x_{(r+1)} \cdot (n-r)}{\sum_{r=0}^{n-1} (n-r)}$$

4. Results

We investigated the role of TikTok’s recommendation algorithm in the propagation of misinformation on the platform. Our primary goal was to understand how users might encounter misinformation through auto-completed query suggestions and search results (RQ1) and to explore how accounts interacting with TikTok videos may impact the amount of misinformation algorithmically recommended to them (RQ2). In the following sections, we present the findings of our four audit experiments, each addressing distinct dimensions of TikTok’s algorithm.

4.1. RQ1: algorithm behaviour when searching for content

Through RQ1, we sought to understand how TikTok’s algorithm could lead to our accounts encountering misinformation through

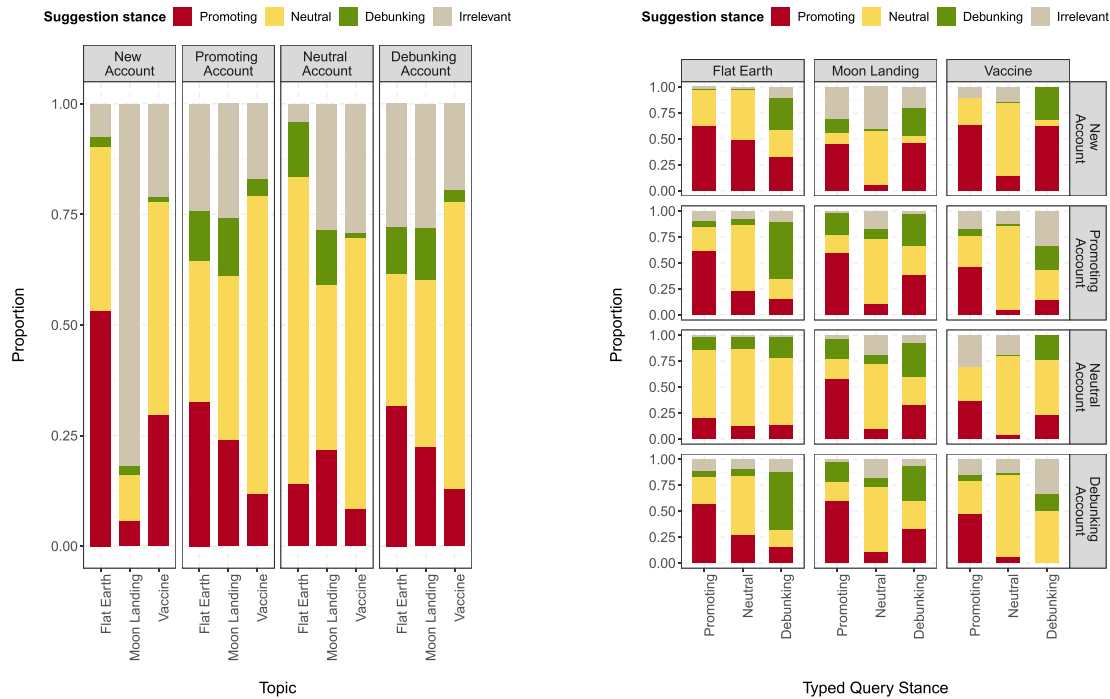


Fig. 5. [RQ1a–Proportion analysis in auto-complete Suggestions] *Left:* Bars depict the proportion of auto-complete suggestions with different stances returned by TikTok, for each topic (bottom) and each account type (top); *Right:* Bars represent the proportion of auto-complete suggestions with different stances returned by TikTok, based on the stance of the query typed into the search bar (bottom), for each topic (top) and account type (right).

auto-complete query suggestions (RQ1a) and search results (RQ1b). To achieve this, we created four different account types: one without prior search history (NEW), and three with specific search histories (PROMOTING, DEBUNKING, and NEUTRAL). We analysed the auto-complete suggestions and search results each account type received while searching for content with three different stances: *promoting*, *debunking*, and *neutral*.

4.1.1. Experiment 1 (RQ1a): when a search query contains a term associated with a misinformation topic, what is the prevalence of misinformation within the auto-complete suggestions?

RQ1a: Proportion Analysis: Overall, our analysis indicates a consistent trend where all account types encounter a significantly higher proportion of auto-complete suggestions *promoting* misinformation compared to those *debunking* it, regardless of the topic (see Fig. 5 (left)). Furthermore, NEW accounts without prior history on the platform received the lowest proportion of *debunking* auto-complete suggestions across all three topics (see Fig. 5 (left)). Additionally, NEW accounts received substantially more *promoting* auto-complete suggestions compared to *debunking* ones, irrespective of the query typed into TikTok's search bar (see Fig. 5 (right)). These findings indicate that NEW users, who have limited familiarity with the platform, likely encounter auto-complete suggestions that support misinformation more frequently than those that challenge it or are neutral in nature.

The proportions of auto-complete suggestion stances were nearly identical across PROMOTING and DEBUNKING accounts, for all topics (see Fig. 5 (left)). All account types entering *debunking* queries across topics received at least some proportion of auto-complete suggestions that promoted misinformation. However, we observed a singular deviation from this trend when DEBUNKING accounts typed queries which aimed to *debunk* vaccine controversies. In this scenario, DEBUNKING accounts did not receive any auto-complete suggestions *promoting* misinformation (see Fig. 5 (right)). Notably, this is the *only* instance where accounts actively looking for content that debunks vaccine-related

misinformation were not presented with any auto-complete suggestions that promoted misinformation instead. Interestingly, NEUTRAL accounts searching for *Flat Earth* content received a significant proportion of *neutral* auto-complete suggestions, regardless of the query stance (see Fig. 5 (right)). However, this trend was not observed for other topics.

RQ1a: Prominence Analysis: The ranking of auto-complete suggestions plays a crucial role in determining the content that users may encounter. Therefore, we analysed the stances of auto-complete suggestions in the top half of each suggestion list (up to four suggestions), when queries of different stances were entered into TikTok's search bar by different account types (see Fig. 6). For example, in Fig. 6, the heatmap in the top-left corner illustrates the prominence scores of auto-complete suggestions *promoting* misinformation when the searched query is also *promoting* in nature, for each topic and account type. Specifically in this heatmap, when a NEW account searches for queries *promoting* the *Flat Earth* theory, the *promoting* auto-complete suggestions from the top half of the list had a prominence score of 0.498. The heatmap on its right illustrates the prominence of *promoting* search results when the query searched is *neutral* in nature, and so on. Overall, our observations indicate that the highest prominence score is consistently obtained when the stance of the search query matches the stance of the auto-complete suggestions.

We found that when typing queries *promoting* misinformation, auto-complete suggestions also *promoting* misinformation were consistently positioned at the top of the suggestion list, irrespective of the account type and misinformation topic. Further, we observed the highest prominence score (0.813) when a NEW account typed *promoting* queries related to *Vaccine Controversies*, indicating that auto-complete suggestions at the top of the list in such cases were predominantly also *promoting* vaccine-related misinformation. Moreover, when the typed query's stance was *neutral*, more auto-complete suggestions *promoting* misinformation were displayed at the top of the list than *debunking* suggestions, across all topics and users. In such cases, the prominence score for auto-complete suggestions promoting misinformation was never zero for any account type. This indicates that users searching for *neutral* queries

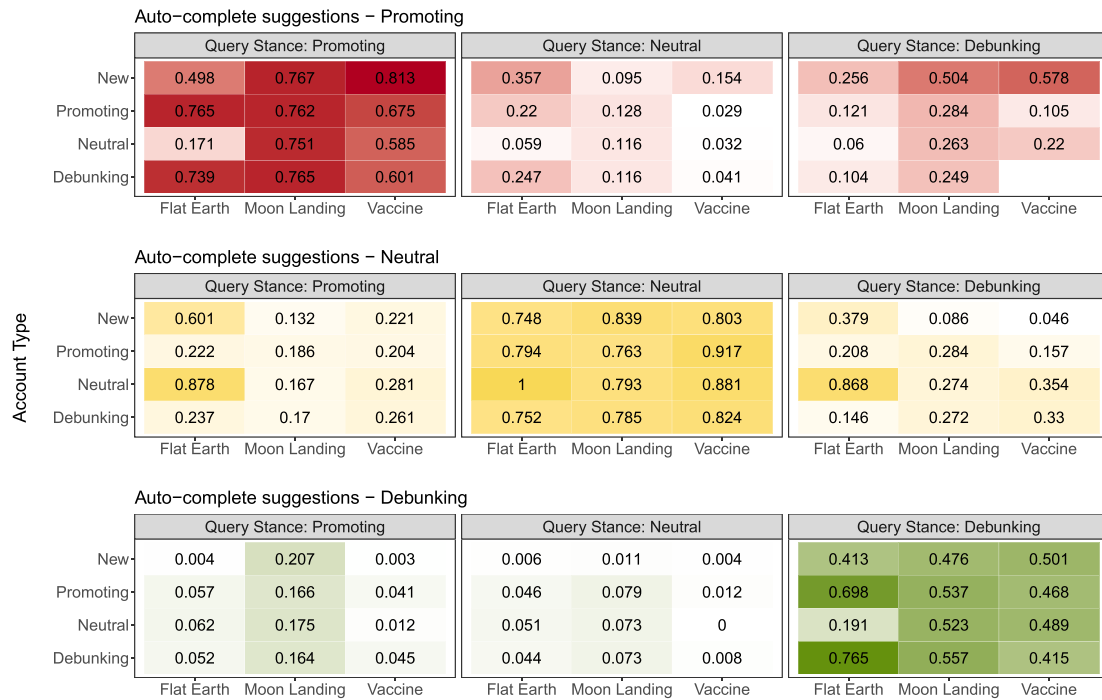


Fig. 6. [RQ1a-Prominence Analysis] The heatmaps display the prominence of each auto-complete suggestion's stance. *Prominence* in this context refers to the visibility and positioning of the auto-complete suggestions. The colours indicate stances of the auto-complete suggestions returned by the platform: red corresponds to *promoting*, green to *debunking*, and yellow to *neutral* stances. Each individual heatmap corresponds to the stance of queries typed on TikTok. Higher colour intensity indicates that suggestions of a specific stance were positioned towards the top of the list, and vice versa.

will likely encounter at least one auto-complete suggestion within the top four that directs them to videos *promoting* misinformation. Further, accounts typing *debunking* queries received auto-complete suggestions *promoting* misinformation, with the highest prominence observed for NEW accounts for the topic of *Vaccine Controversies* (0.578).

4.1.2. Experiment 2 (RQ1b): when a search query contains a term associated with a misinformation topic, what is the prevalence of misinformation within the search results?

RQ1b: Proportion Analysis: Our investigation reveals that the proportion of misinformation recommended by the platform differed greatly based on the topic. All account types searching queries related to the topic *Flat Earth* encountered the highest proportion of results *promoting* misinformation, followed by the topic *Moon Landing*, with searches on *Vaccine Controversy* receiving the fewest results *promoting* misinformation. This is illustrated in the staircase pattern observed in Fig. 7 (left). Moreover, for each topic, NEW accounts consistently received the highest proportion of search results *promoting* misinformation when compared to other account types. In addition to search results *promoting* misinformation, all account types also encountered search results that debunked misinformation. However, queries related to *Vaccine Controversies* received the smallest proportion of search results *debunking* misinformation, compared to the other two topics, across account types.

We observed that NEW accounts being recommended more misinformation occurred irrespective of the stance of the query they searched, particularly for the topic *Flat Earth*. Further, NEW accounts witnessed a higher proportion of misinformation *promoting* materials when their search query was *neutral* than when it was *debunking* in nature. Conversely, for the topic *Vaccine Controversies*, all accounts searching for *neutral* queries received some of the smallest proportions of results that were *promoting* misinformation. This observation becomes even more stark when we specifically focus on PROMOTING accounts searching for *neutral* queries on *Vaccine Controversies*. In such scenarios,

TikTok's algorithm returned no misinformation whatsoever, with the results containing only *neutral* or *irrelevant* videos. Further, the algorithm's sensitivity to query stance was most evident in the *Flat Earth* (and to a lesser extent *Moon Landing*), where *debunking* queries returned a higher proportion of *debunking* results compared to *promoting* and *neutral* queries. This pattern was not observed for the *Vaccine Controversies*, where *debunking* queries yielded similar proportions of *neutral* and *debunking* search results across user types. Finally, the algorithm served a considerable proportion of *neutral* content to NEUTRAL accounts searching *neutral* queries. However, for the topic *Flat Earth*, only a very limited volume of results were *neutral* in stance.

RQ1b: Prominence Analysis: In this analysis, our aim was to examine the visibility and positioning of misinformation videos within TikTok's search results under different scenarios. To gauge this, we calculated a prominence score based on the positioning of misinformation videos in the top 10 search results when a specific query is entered. For example, in Fig. 8, the heatmap in the top-left corner illustrates the prominence scores of search results *promoting* misinformation when the searched query is also *promoting* in nature, for each topic and account type. Specifically in this heatmap, when a NEW account searches for queries *promoting* the *Flat Earth* theory, the search results returned by TikTok contain videos at the top which also promote this misinformation, with prominence score of 0.808. The heatmap on its right illustrates the prominence of *promoting* search results when the query searched is *neutral* in nature, and so on. Overall, our observations indicate that the highest prominence score is consistently obtained when the stance of the search query matches the stance of the search results.

When examining the prominence scores across different account types, a clear trend emerges. When NEW accounts searched for *promoting*, *neutral*, and *debunking* queries across different topics, search results *promoting* misinformation had the highest prominence scores in 7 out of 9 cases, compared to *debunking* search results (see Fig. 8). This suggests that, regardless of the topic or query stance, NEW accounts tend to have *promoting* search results prominently displayed at the top of the list more

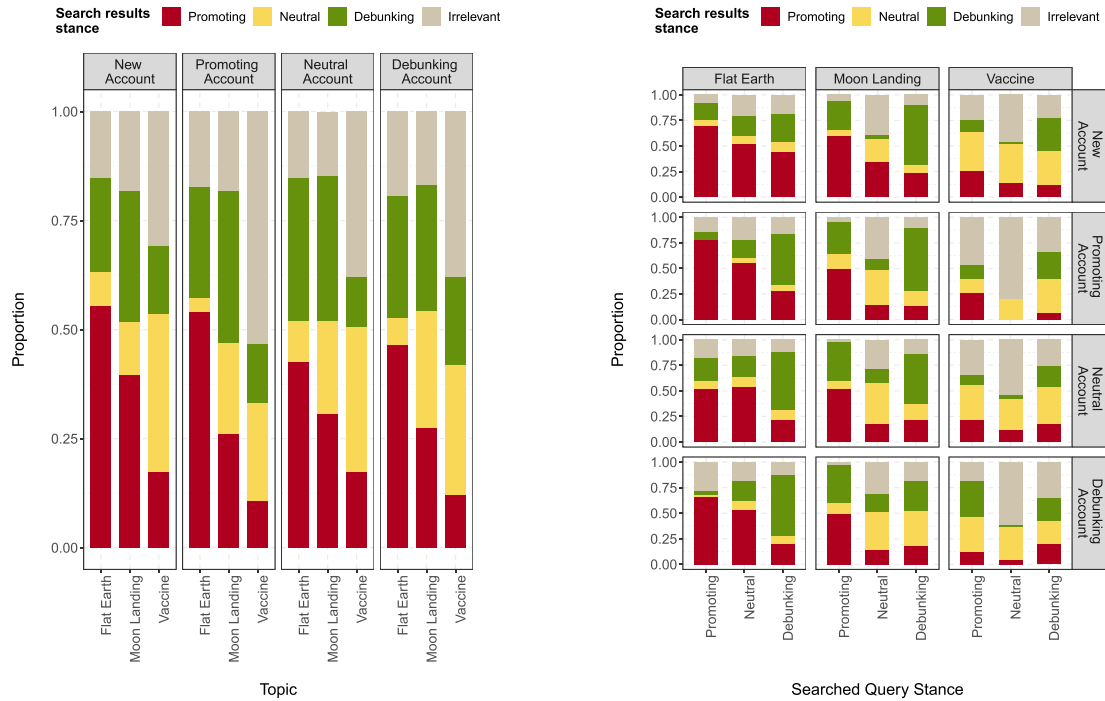


Fig. 7. [RQ1b-Proportion analysis in search Results] *Left:* Bars depict the proportion of search results with different stances returned by TikTok, for each topic (bottom) and each account type (top); *Right:* Bars represent the proportion of search results with different stances returned by TikTok, based on the stance of the query searched (bottom), for each topic (top) and account type (right).

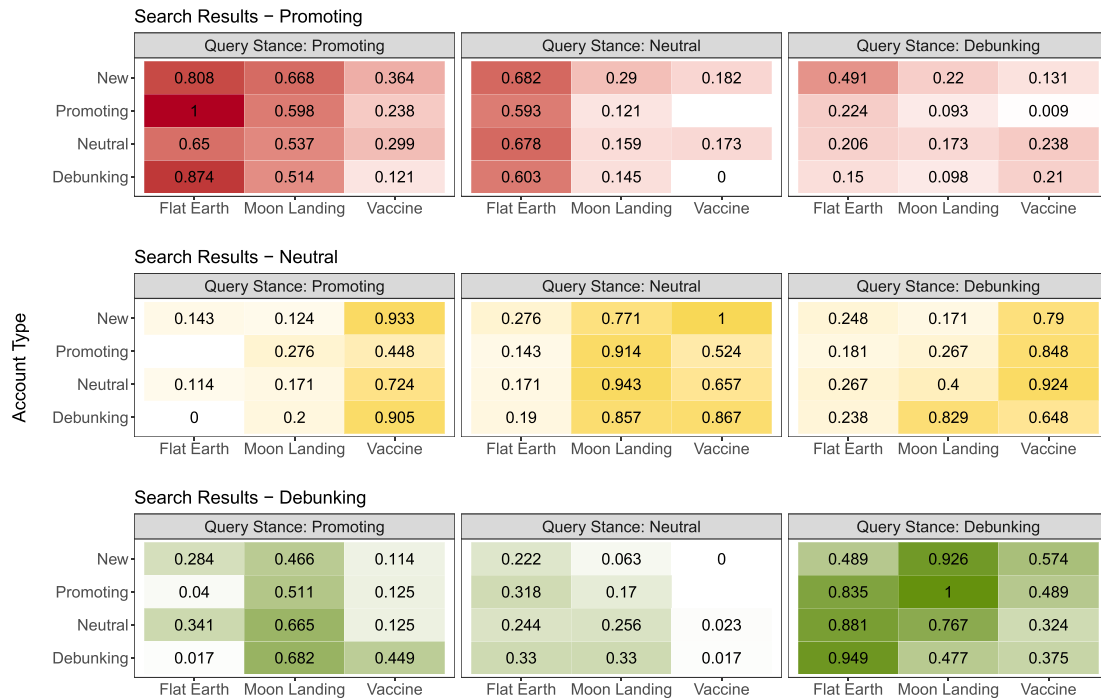


Fig. 8. [RQ1b-Prominence Analysis] The heatmaps display the prominence of each search result stance. *Prominence* in this context refers to the visibility and positioning of the search results. The colours indicate stances of the search results returned by the platform: red corresponds to *promoting*, green to *debunking*, and yellow to *neutral* stances. Each individual heatmap corresponds to the stance of queries searched on TikTok. Higher colour intensity indicates that search results of a specific stance were positioned towards the top of the results page, and vice versa.

frequently than other account types. Even when the search queries were *neutral* in nature, TikTok's search results more prominently displayed videos that were *promoting* misinformation, compared to those with a *debunking* or *neutral* stance, for topics such as *Flat Earth*. Conversely, we note that *debunking* results were more prominent when all accounts

searched for queries *promoting* misinformation related to the topic *Moon Landing*, compared to the other two topics.

Moreover, when comparing the prominence of videos *promoting* vaccine-related misinformation between *NEUTRAL* and *PROMOTING* account types, we observe that *NEUTRAL* accounts consistently see higher

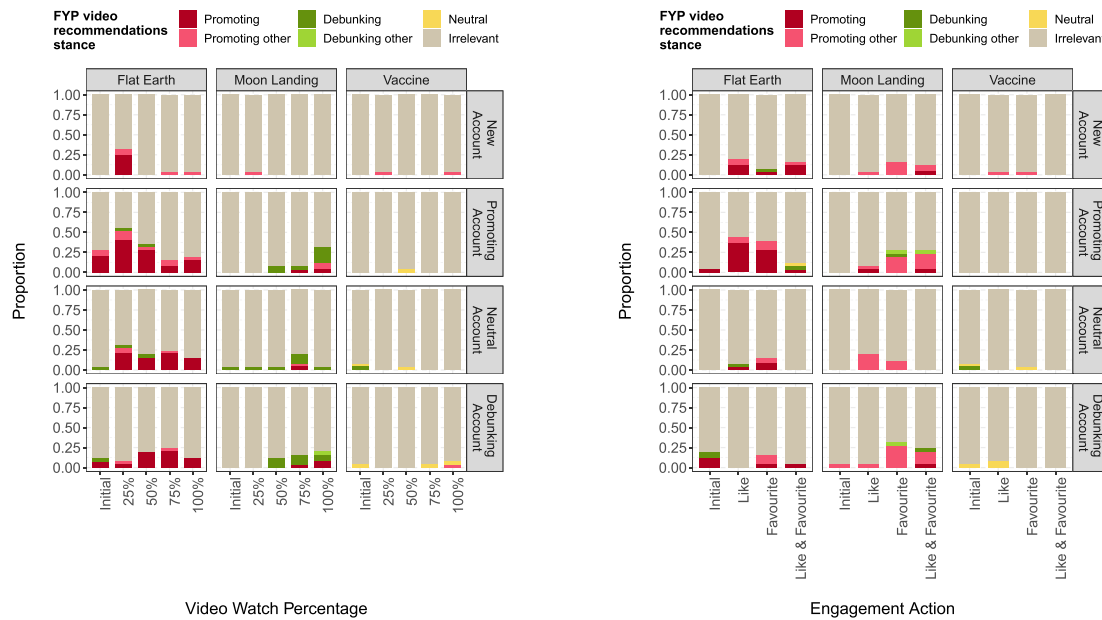


Fig. 9. [RQ2a (Left) & RQ2b (Right) – Proportion Analysis] Plots depict the proportions of video stances on the For You Page (FYP) after engaging with *promoting* videos in different ways. *Left*: plot represents the outcomes of Experiment 3 analysing the impact of different watch percentages on FYP content; *Right*: plot represents the outcomes of Experiment 4 analysing the impact of different engagement actions on FYP content. *Initial* represents the stance composition of the FYP before these actions were undertaken.

prominence scores for such misinformative videos. This suggests that NEUTRAL accounts, which are simply seeking information, are more likely to have videos *promoting* misinformation prominently displayed in their search results, compared to PROMOTING accounts which actively search for misinformation.

4.2. RQ2: algorithm behaviour after engagement with recommended videos

Our second research question aims to understand how watching recommended videos and engaging with them in different ways influence the prevalence of misinformation recommended by TikTok's algorithm. Similar to the SEARCH Experiments, our two ENGAGEMENT Experiments also encompass four different account types (NEW, PROMOTING, DEBUNKING, NEUTRAL) and three misinformation topics (*Moon Landing*, *Flat Earth*, and *Vaccine Controversies*). To answer RQ2, we present our findings for the watching (S3–S6) and engagement (S7–S9) scenarios from Table 1. We collected the initial 25 videos recommended by TikTok's algorithm on our accounts' For You Pages (FYPs) twice—both before and after the accounts executed the watching or engagement actions outlined in their respective scenarios. In the following sections, we present our findings from *proportion* and *prominence* analyses of videos recommended by TikTok, with a specific emphasis on videos *promoting* and *debunking* misinformation.

4.2.1. Experiment 3 (RQ2a): how does a user's watch time for misinformation videos influence the content subsequently displayed on their for you page?

RQ2a: Proportion Analysis: More than 50% of the videos on the For You Page (FYP) were labelled 'irrelevant.' This signifies that the majority of FYP videos were not related to the three topics under consideration, and also did not engage in discussions *promoting* or *debunking* other controversial topics, which we labelled as *promoting-other* and *debunking-other* during our labelling process. This aligns with the intended purpose of a For You Page—to showcase diverse content to users.

However, after our accounts watched videos *promoting* misinformation, we observed distinct changes in the prevalence of misinformation on their For You Pages (FYPs). Specifically, we observed that TikTok's

algorithm moderated the prevalence of misinformation differently for different topics under investigation (see Fig. 9 (left)). For the *Flat Earth* topic, the algorithm appeared to be highly sensitive to the stance of the videos watched by accounts, with more videos *promoting* *Flat Earth* misinformation as well as misinformation on other topics (*promoting-other*) infiltrating the refreshed FYPs. However, for the *Moon Landing* topic, despite accounts having watched videos *promoting* misinformation, the refreshed FYP witnessed an increased proportion of videos *debunking* misinformation, but those *promoting* it remained relatively similar in proportion to the *initial* FYP. In contrast to both *Flat Earth* and *Moon Landing*, the algorithm behaved differently yet again for the *Vaccine* topic. On the refreshed FYP, the proportions of video stances remained relatively similar, with the algorithm appearing to be almost indifferent to accounts watching videos *promoting* vaccine-related misinformation.

It is noteworthy that the highest changes in proportions of video stances on the FYP were observed for NEW accounts which had watched *Flat Earth* videos. These accounts not only witnessed a notable surge in the proportion of videos *promoting* misinformation related to *Flat Earth*, but videos *promoting* misinformation on other topics also found their way into these FYPs (*promoting-other*) (see Fig. 9 (left)). Surprisingly, the rise in the proportion of *promoting* and *promoting-other* videos did not linearly correlate with watch percentages: accounts watching a *larger* percentage of *promoting* videos did not result in a proportional increase in the prevalence of *promoting* videos in the FYP, compared to accounts watching a *smaller* percentage of such videos.

RQ2a: Prominence Analysis: Overall, the prominence of videos *promoting* misinformation on the FYP did not exhibit a linear correlation with watch percentages. It was sufficient for NEW, PROMOTING, and NEUTRAL accounts to watch only 25% of a *promoting* video to have videos *promoting* misinformation be more prominently displayed on their FYPs, compared to watching longer durations (see Fig. 10). Interestingly, even with a comparably lower occurrence of recommendations endorsing the *Vaccine Controversy*, as evidenced by the nearly vacant heatmap cells for this topic, *New* accounts witnessed the highest prominence score for this topic by watching videos *promoting* vaccine-related misinformation for a mere 25% of their duration. However, for the topic *Moon Landing*, the prominence of *promoting* recommendations

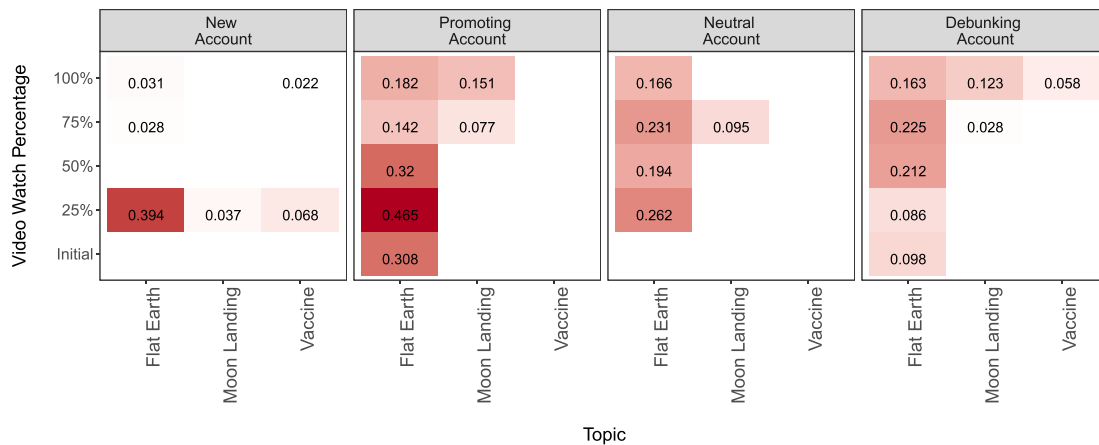


Fig. 10. [RQ2a–Prominence Analysis] The heatmaps display the prominence of videos promoting misinformation on the For You Page (FYP). *Prominence* in this context refers to the visibility and positioning of the video within the top 25 recommendations on the FYP. These values are in response to different video watch percentages (25%, 50%, 75%, 100%) for each account type and topic. *Initial* represents the stance composition of the FYP before these watching actions were undertaken. Higher colour intensity indicates that search results of a specific stance were positioned towards the top of the FYP, and vice versa.

increased as the watch percentage rose to 100% in PROMOTING and DEBUNKING accounts. This suggests that for accounts that have searched either *promoting* or *debunking* misinformation related to *Moon Landing*, the algorithm appears to require longer durations to cause subsequent changes in their FYPs.

The influence of having a *debunking* search history is especially apparent for the topic *Flat Earth*, as DEBUNKING accounts exhibited a higher prominence score for *promoting* recommendations only after watching just half of the video (50%), compared to the other three account types.

4.2.2. Experiment 4 (RQ2b): how does user engagement with videos, such as liking and favouriting, influence the content displayed on the for you page?

RQ2b: Proportion Analysis: Out of all the FYP videos collected across all actions and account types for Experiment 4, 14.75% were *promoting* misinformation. Notably, after accounts undertook engagement actions, there was a notable increase in *promoting* content being recommended for the *Moon Landing* topic, which was not observed in Experiment 3 when accounts watched videos. Interestingly, for *Vaccine Controversies*, the algorithm recommended more *neutral* content post-engagement, a finding also consistent with results of Experiment 3. Moreover, for *Vaccine Controversies*, NEW accounts received some recommendations *promoting* misinformation on other topics (*promoting-other*), while the remaining account types did not receive any recommendations *promoting* misinformation whatsoever.

Further, *promoting* videos were recommended more to accounts after they *Liked* and *Favourited* *promoting* videos for the *Flat Earth* topic. However, for *Moon Landing*, more *promoting* videos were recommended when accounts *Favourited*, and *Liked* & *Favourited* videos. Notably, engaging with *promoting* videos related to the *Moon Landing* topic triggered the appearance of misinformation-promoting videos on other topics, serving as a gateway to other problematic content. Finally, DEBUNKING accounts exhibited an increase in promoting recommendations compared to their initial FYP composition when engaging with *Moon Landing*-related content, particularly under favouriting actions, while this pattern was weaker for *Flat Earth* and not observed for *Vaccine Controversies*.

RQ2b: Prominence Analysis: We did not observe a discernible trend or correlation between the type of engagement action and subsequent changes in the prominence of videos with different stances. However, it was notable that *Promoting* recommendations became more prominently positioned for the *Flat Earth* topic, particularly when PROMOTING

accounts performed *Like* and *Favourite* actions (see Fig. 11). Further, for the *Moon Landing* topic, even accounts with a DEBUNKING search history saw videos *promoting* misinformation achieve greater prominence on their FYPs, after performing various engagement actions.

5. Discussion

Short-form videos have emerged as a prominent channel for disseminating misinformation, raising concerns about the role of platforms hosting these videos in regulating information distribution. Specifically, it has been alleged that these platforms sometimes prioritise result relevance without adequately considering the accuracy and credibility of the information presented. In response to multiple reports highlighting the prevalence of misinformation on short video applications such as TikTok (Bakan, 2023; O'Connor, 2021; Hsu, 2022; Tucker, 2022), we conducted an empirical audit of TikTok's recommendation algorithm. Our aim was to systematically assess the extent of misinformative content within the search auto-completion suggestions, as well as the searched and recommended videos displayed by the platform.

5.1. Prevalence of misinformation when searching for content on TikTok (RQ1)

Auto-complete suggestions which are biased or harmful can reinforce existing stereotypes and misconceptions, and may even subtly encourage users to adopt undesirable behaviours (Olteanu et al., 2020). In an ideal world, users would first assess each of the suggestions before clicking on them. However, the curiosity or shock value associated with misinformation related suggestions can lead users to click on them. While it can be assumed that only users who are already interested in conspiratorial content search for such content, our results show that searching a *neutral* query (e.g., 'moon landing', 'vaccine') can also lead to seeing auto-complete suggestions and search results *promoting* misinformation (see Fig. 5 (right) and Fig. 7 (right)). Our findings also align with a report from the NewsGuard which stated that neutral phrases such as "climate change" or "COVID vaccine" typed in TikTok's search bar resulted into misleading phrases being suggested, such as "climate change doesn't exist" and "COVID vaccine HIV" (Brewster et al., 2022).

Encountering conspiratorial information while casually exploring a topic can lead to the adoption of these conspiratorial beliefs (Ecker et al., 2022). This becomes especially concerning when the user has limited knowledge about the given topic, prompting them to search for information on it in the first place. In fact, Sallam et al. (2020) argued that lower domain knowledge about a specific topic is a contributing factor to individuals believing in conspiracy theories. Notably, our results show that

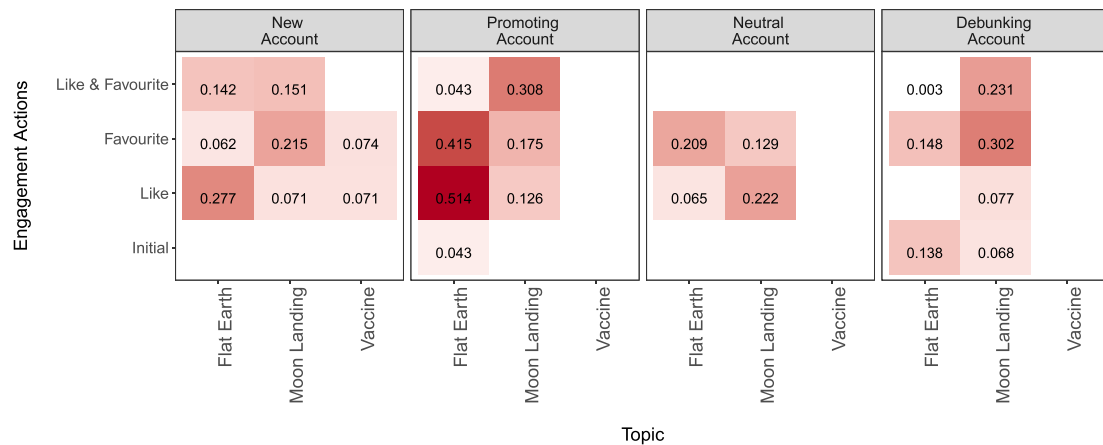


Fig. 11. [RQ2b–Prominence Analysis] The heatmaps display the prominence of videos promoting misinformation on the For You Page (FYP). *Prominence* in this context refers to the visibility and positioning of the video within the top 25 recommendations on the FYP. These values are in response to different engagement actions (*Like*, *Favourite*, and *Like & Favourite*) for each account type and topic. *Initial* represents the stance composition of the FYP before these actions were undertaken. Higher colour intensity indicates that search results of a specific stance were positioned towards the top of the FYP, and vice versa.

TikTok’s algorithm prominently positioned problematic auto-complete suggestions and search results at the top of their respective lists, as illustrated in Figs. 6 and 8. Interestingly, they are ranked even higher than *debunking* auto-complete suggestions and search results, exacerbating the issue further. It is not uncommon for users to click on the top-listed results, as they implicitly trust the particular search engine, which can significantly impact the user’s opinion on a given matter (Kulshrestha et al., 2017). Additionally, this may feed into a toxic loop—if users consistently choose such suggestions, similar suggestions are also more likely to appear at the top of the list repeatedly, which the users would then be even more exposed to.

Unfortunately, a user casually searching for information related to a conspiratorial topic, might inadvertently go down a rabbit hole simply because problematic content is prominently visible and positioned. Our findings show that NEW accounts appear to be the most vulnerable account type, with a higher likelihood of falling into this rabbit hole and being the most susceptible to algorithmically recommended misinformation. NEW accounts received the smallest proportion of auto-complete suggestions that debunked misinformation, compared to other account types, and also received the highest proportion of search results promoting misinformation. For example, for such accounts, the proportion of search results promoting misinformation increased by 35% when they merely searched a neutral query related to the topic *Flat Earth* (refer to Fig. 5). While the primary objective of social media platforms is to cater to user needs, it is crucial to acknowledge their concurrent aim of maximising revenue. This often leads platforms to prioritise content that captivates users and sustains their engagement. The popularity of such captivating content can result in its increased visibility, potentially influencing the content that appears as recommendations. Previous work has shown that YouTube actively hosts and recommends videos violating its own policies on political and medical misinformation, hate speech, and inappropriate content (McCrosky and Geurkink, 2021) because of the level of engagement such videos garner. The stimulating nature of misinformation content contributes to the very prevalence of such content, potentially explaining why NEW accounts are frequently exposed to misinformation by TikTok’s algorithm.

5.2. Impact of video engagement on misinformation presented on TikTok (RQ2)

Users on recommendation platforms encounter content curated by the platform, and if this content contains misinformation, its consumption may contribute to the amplification of such material within the platform (Tomlein et al., 2021). This amplification in turn can lead to

the creation of a filter bubble effect, further entrenching users deeper into a state of misinformed beliefs. One of the objectives of our audit investigation was to understand the influence of direct interactions with videos presented as recommendations on TikTok’s For You Page. To facilitate this analysis, we categorised account actions into two distinct groups: *initial* actions, such as watching videos, and *secondary* actions, such as liking and favouriting videos.

We found that once engagement with misinformation promoting videos begins, it can influence the recommendations provided in the FYP. As illustrated in Fig. 9-left, for instance, PROMOTING accounts received twice as many misinformation *promoting* recommendations by merely watching 25% of these types of videos. Engaging with misinformation *promoting* videos not only generates more problematic recommendations but can also create a filter bubble effect, restricting users from encountering views that differ from their pre-existing ones, potentially decreasing the diversity of content on their FYP. Spinelli and Crovella (2020) also show how YouTube’s recommendations often “lead users away” from reliable information sources. Their study revealed a tendency for the platform to gradually guide users towards video channels that promote extreme and unscientific viewpoints.

While TikTok’s algorithm aims to cater to user preferences by providing relevant content, a broader ethical question arises: Should a platform deliver potentially harmful content to users solely based on their expressed interest in it? Given TikTok’s widespread use, the platform bears a social responsibility to ensure that users receive accurate information. Our findings indicate that TikTok seems to adopt such practices, but only for certain topics. For example, the topic of *Flat Earth* received a higher proportion of recommendations that promote misinformation, while considerably fewer were suggested for the topic of *Vaccine Controversies*. TikTok has explicitly mentioned in its community guidelines that it does not permit medical misinformation regarding vaccines or COVID-19 (Keenan, 2022). This phenomenon can also be seen in YouTube’s recommendation algorithm: Hussein et al. (2020) noted that users watching videos on *Vaccine Controversies* received a negligible proportion of videos promoting “vaccine hesitancy”. Similarly, Papadamou et al. (2020) affirm that YouTube tends to suggest more pseudo-scientific content related to traditional topics like *Flat Earth*, as opposed to emerging ones. It is apparent that akin to YouTube, TikTok also selectively moderates topics that are sensitive or more controversial, exercising control over their visibility on the platform.

In addition, even when misinformative content on certain topics may be restricted, the platform often recommends misinformation on other unrelated topics, in response to user actions. In Fig. 9 (right), it is evident that NEW accounts received recommendations that promote

misinformation on different topics, when actions were taken on *promoting* content related to the topic *Vaccine Controversies*. Furthermore, all account types that have interacted with *promoting* videos related to the topic *Moon Landing* received a significant amount of misinformation *promoting* other topics. One potential reason for this could be the presence of hashtags in these videos, such as #conspiracy, #conspiracytheories, and #conspiracytiktok. These hashtags are also present in other misinformation videos, and it appears that the algorithm is recommending videos based on shared hashtags. Therefore, a critical question arises: Is controlling a limited number of topics sufficient to effectively combat misinformation within these platforms? Our results show that TikTok is effectively restricting misinformation related to *Vaccine Controversies*; however, it does not achieve the same level of success in addressing misinformation concerning the other topics (*Flat Earth* and *Moon Landing*). Furthermore, despite TikTok's efforts to limit misinformation on *Vaccine Controversies*, their approach has inadvertently facilitated exposure to other types of misinformation, as shown in our findings. Therefore, platforms aiming to reduce misinformation in their ecosystem should implement broader strategies to comprehensively combat misinformation. This is crucial to ensure that misinformation on a given topic is not simply replaced by other forms of misinformation in search results and recommendations provided to users.

5.3. Algorithm transparency: ensuring trust and accountability

Recent studies have repeatedly showcased that video is more convincing than other forms of modalities (Sundar et al., 2021; Appiah, 2006; Wittenberg et al., 2021). These studies underscore that, unlike text formats which demand additional effort from users to comprehend the author's message, the inherent visual aspect of the video modality directly appeals to our senses. In particular, short-form videos have changed the way people interact with the world, get entertained, and consume news. The short format makes it easier for users to consume information quickly and effortlessly, establishing it as an increasingly popular source for news consumption (Tomasik and Matsa, 2025; Sillito, 2022).

Platforms such as TikTok have asserted that they are actively taking measures to combat misinformation. Some of these measures include partnering with fact-checkers and displaying disclaimer banners to alert end-users.⁸ In addition, over 102 million videos were removed from the platform for violating TikTok's community guidelines, in the first quarter of 2022 (MENAT, 2022). However, only 1% of these videos were removed for violating the platform's "integrity and authenticity" guidelines, which include "harmful misinformation" (Singh, 2022). This is supported by our findings which clearly indicate that TikTok's recommendation algorithm pushes a significant volume of misinformation onto users. TikTok mentions their recommendations are based on a number of factors such as user interactions and video information.⁹ Nevertheless, this information also fails to clarify why new users with no prior interactions on the platform were repeatedly exposed to misinformation. The exposure of users to algorithmically curated misinformation through short videos is not an issue unique to TikTok. Other platforms with similar content, such as Instagram Reels and YouTube Shorts, also face accusations of facilitating the spread of harmful content (Sardarizadeh, 2019; YouTube May Have Misinformation Blind Spots, 2022; de Luna, 2022; Elise Thomas, 2022; Clara Martiny and Cooper, 2022). In addition to the short-form video format, these platforms share other similarities. They offer users the ability to search for content, provide auto-complete suggestions during searches, and

continuously deliver endless recommendations in the feed until the user disengages. Similar to the measures implemented by TikTok, these platforms have incorporated methods such as removing harmful content, flagging misinformation materials and also using machine learning techniques to automatically identify harmful content (Instagram, 2026; Mohan, 2021). However, these efforts have not been able to completely stop the spread of misinformation within these platforms (ARMSTRONG, 2023; Clara Martiny and Cooper, 2022).

Therefore, we argue that it is essential to subject almost any algorithm to critical examination. While algorithms of platforms such as TikTok and YouTube are primarily designed to cater to user needs, it is worth noting that most algorithmic systems often pursue multiple objectives simultaneously. It is crucial to understand that algorithms can be "rigged" in ways that raise ethical concerns. Therefore, we argue that public scrutiny of algorithms is necessary. Among the various challenges of studying misinformation on short-video platforms, an important one is the lack of API support provided to researchers investigating misinformation on these platforms. While API systems for accessing these platforms' data are essential for large-scale data-driven research, their current state is deemed insufficient (Venkatagiri et al., 2023). Platforms should actively share data with independent auditors, allowing them to empirically present evidence of the platforms' effectiveness and identify any shortcomings. This is particularly vital as algorithmic transparency gains significance in safeguarding the integrity of information dissemination, particularly on platforms with more "convincing" content such as those with short videos. This auditing process, conducted in parallel with the interventions implemented by the platforms themselves and more general policy changes, will contribute to transparency and accountability, and foster public trust.

5.4. Limitations and future work

Our study has several limitations. First, we examined a fixed number of user actions and misinformation topics. There is a great variety of different user actions, such as following the content creator and commenting on videos, that can also influence a complex recommendation algorithm like the one used by TikTok. Future work should investigate how recommendations are affected when users engage in more complex behaviours, such as commenting on and sharing problematic content. This will help to understand the consequences of using a wider array of platform interactions and how they influence the spread of misinformation.

Second, to capture the full spectrum of misinformation on TikTok, it would be interesting to examine how the recommendations differ for other established and emerging misinformation topics. Additionally, it would be insightful to investigate which types of topics receive more or less visibility.

Third, our study was limited to only one short video platform. Building on our approach, future studies can be conducted on similar platforms, such as Instagram Reels and YouTube Shorts, to further understand the role that these recommendation algorithms play with regard to the dissemination of misinformative content. In addition, future studies can utilise a data donations framework to gather real user data and analyse it to better understand the behaviour of recommendations due to the users interacting with problematic content.

Fourth, we did not account for the popularity of the content when assessing the prevalence of misinformation content on TikTok. Future studies can analyse various social engagement cues such as number of likes, comments and shares which contribute to the popularity of videos, to gain a deeper understanding of the issue.

Finally, we did not consider the role of user demographics, such as location and age. By keeping demographic factors static when creating different accounts, we minimized potential biases arising from demographic variations (Boeker and Uрман, 2022). This allowed for an

⁸ <https://www.tiktok.com/transparency/en-us/combating-misinformation/>.

⁹ <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you>.

in-depth examination of the impact of user actions and account histories on the prevalence of misinformation recommended by TikTok's algorithm. However, we acknowledge the potential impact of these factors on the content. As such, future studies should conduct a more in-depth investigation across different demographics in order to provide a more nuanced view of how misinformation is recommended to different types of users.

6. Conclusion

In this study, we conducted four audit experiments on TikTok to empirically understand its recommendation algorithm's behaviour, when accounts search for specific types of content and engage with video recommendations in different ways. These experiments were conducted on three popular topics linked to misinformation and conspiracy theories: *Flat Earth*, *Moon Landing*, and *Vaccine Controversies*. The impact of these actions was measured across four different account types, including new accounts and those with varying search histories. Alarming, we found that new users are susceptible to algorithmically recommended misinformation content. Furthermore, our findings reveal how different user actions can lead to more problematic content being recommended by TikTok. Interestingly, we show that TikTok is capable of stifling misinformation content for specific topics such as *Vaccine Controversies*, only to replace it with misinformation on unrelated topics. In this work, we offer key insights into the factors influencing the prevalence and placement of misinformation on short video platforms like TikTok, enhancing our understanding of content recommendation algorithms in the era of rich media consumption. We contend that such algorithmic audits are crucial for upholding the transparency and credibility of recommendation platforms, especially those that function as sources of information for their end-users.

CRediT authorship contribution statement

Suwani Gunasekara: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Saumya Pareek:** Writing – review & editing, Formal analysis. **Ryan M. Kelly:** Writing – review & editing, Supervision. **Vassilis Kostakos:** Writing – review & editing, Supervision. **Jorge Goncalves:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Prominence score calculation example

The variable r indicates the position (rank) of a video within the list, starting from 0. For example, a video which was listed as first in a specific search result will have the rank of zero, one which was listed as the second will have rank of one, and so on.

Take two search result lists which contain 5 videos for a given topic t . If a particular position contains a video which promotes misinformation, it is denoted as 1; otherwise it is denoted as 0. $list1 = [1, 0, 0, 0, 0]$, $list2 = [0, 0, 0, 0, 1]$. In these two search result lists, $list1$ has a misinformation promoting video at the beginning of the list ($r = 0$) and in $list2$, there is a misinformation promoting video at the end of the list ($r = 4$). x_r indicates the boolean value of that particular r position. For example, in $list1$, $x_0 = 1$ and $x_1 = 0$. $(n - r)$ will be the weight given to each video, with n being the total number of videos in that list. Elements positioned towards the top of the list will receive more weight than elements positioned at the bottom.

Prominence score of $list1$

$$\begin{aligned} &= \frac{x_0.(5-0) + x_1.(5-1) + x_2.(5-2) + x_3.(5-3) + x_4.(5-4)}{(5-0).(5-1).(5-2).(5-3).(5-4)} \\ &= \frac{1.(5-0) + 0.(5-1) + 0.(5-2) + 0.(5-3) + 0.(5-4)}{(5-0).(5-1).(5-2).(5-3).(5-4)} \\ &= \frac{5}{120} \\ &= 0.0416 \end{aligned}$$

Prominence score of $list2$

$$\begin{aligned} &= \frac{x_0.(5-0) + x_1.(5-1) + x_2.(5-2) + x_3.(5-3) + x_4.(5-4)}{(5-0).(5-1).(5-2).(5-3).(5-4)} \\ &= \frac{0.(5-0) + 0.(5-1) + 0.(5-2) + 0.(5-3) + 1.(5-4)}{(5-0).(5-1).(5-2).(5-3).(5-4)} \\ &= \frac{1}{120} \\ &= 0.0083 \end{aligned}$$

$list1$ has a higher prominence score as it has a misinformation promoting video search result more towards the top of the list compared to $list2$.

Appendix B. Examples and stance labels for each misinformation topic

Please note that the cells in the last column in [Tables B.3 and B.4](#) contain the explanations generated by GPT-4, which have not been edited by the researchers. The model has emphasised certain phrases using bold text, which is shown by two asterisks following the Markdown syntax.

B.1. Autocomplete suggestions

Table B.3

Sample of auto-complete suggestions generated for the topic *Moon Landing* along with their relevant labels and model output generated.

Auto-complete suggestions	Label	ChatGPT output
the moon landing was fake	Promoting	***the moon landing was fake - Promoting** This query directly claims that the moon landing was fake, promoting misinformation."
stanley kubrick conspiracy moon landing	Promoting	***stanley kubrick conspiracy moon landing - Promoting** This query seeks content that proves or supports misinformation related to the topic Moon Landing."
moon landing interview	Neutral	***moon landing interview - Neutral** This query simply seeks interviews related to the moon landing without indicating whether the content supports or refutes misinformation."
first man on the moon	Neutral	***first man on the moon - Neutral** This query seeks information about the first person to land on the moon and doesn't imply any bias towards promoting or debunking misinformation."
moon landing conspiracy debunked	Debunking	***moon landing conspiracy debunked - Debunking** This query seeks content that debunks the conspiracy theories regarding the moon landing, thus refuting misinformation."
why the moon landing was not fake	Debunking	***why the moon landing was not fake - Debunking** This query is looking for content that explains why the moon landing was genuine, thereby debunking the idea that it was faked."

B.2. Videos

Table B.4

Video transcript, label and the generated output from the model for a video on the topic *Flat Earth* when labelling videos related to the topic *Moon Landing*.

Video Transcript
Start by taking off this dome here. That was so impossible. Alright, starting with the base, I have three magnets and as you see, self is pointed upwards which works out perfectly for a compass. Now, I've tried four magnets around here, but a lot of people think of Earth as a ring magnet and some people think of it as a magnetic pump. Well, it actually could be both. What we'll do is, set you up here, if we get started on this. We'll zoom in so you can, hopefully you can see better. We're going to go a little darker. We can turn on some lights and we'll go and try it out. And then we'll let it go. Alright, the first thing we'll do is, all I have to do is salt water to the center, being careful not to worry about confidence. Now, I've seen something like this from a YouTuber named everybody's eye. Had not been doing it flattered and he doesn't even believe in flattered. And, you know, it was two years ago at the time I didn't either. But now, it seems the more and more I hear about the globe Earth, the more ridiculous it sounds. Alright, now I'm going to hook up my terminals to the negative, we'll go to the outer ring, and the positive to the center. Then with my battery charger, I'm going to turn this on at 10 amps, say, that's 12 volts, of course. Oh, and so we can see what's going on here. I'm going to add a little bit of pepper. Alright, I'll let this run a little bit and you can see how the currents work. By turning this, it'll actually, it seems like it times it. So, the water will actually move a little bit differently. But anyway, take a look at it. Let me know what you think in the comments and thanks for watching. Thank you. Thank you.
Label
Promoting-Other
ChatGPT Output
Label: 4. Promoting-Other This transcript is not directly related to the Moon Landing but is discussing magnetism and flat Earth theories. The speaker is demonstrating an experiment using magnets, water, and various substances. He implies that the more he learns about the theory of globe Earth, the more absurd it sounds to him, indicating a promotion of misinformation regarding the shape of the Earth. This falls into 'Promoting-Other' as it's promoting misinformation on another topic, not explicitly related to the Moon Landing.

Data availability

An OSF link to the data used is provided in the manuscript.

References

- Abualsaud, M., Smucker, M.D., 2019. Exposure and order effects of misinformation on health search decisions. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Rome.
- Ai, O., 2022. Introducing whisper. <https://openai.com/research/whisper>.
- Akbar, A., Caton, S., Bierig, R., 2023. Personalised filter bias with Google and DuckDuckGo: an exploratory study. In: Longo, L., O'reilly, R. (Eds.), *Artificial Intelligence and Cognitive Science*. Springer Nature Switzerland, Cham, pp. 502–513.
- Alex Olshansky, B., 2018. Conspiracy theorizing and religious motivated reasoning: why the earth 'must' be flat. <https://ttu-ir.tdl.org/handle/2346/82666> (Accessed 13 October 2023).
- Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Bermeo, J.D., Korobeynikova, M., Gilardi, F., 2023. Open-source large language models outperform crowd workers and approach ChatGPT in text-annotation tasks. *arXiv preprint arXiv:2307.02179*.
- Allem, J.-P., 2020. Social media fuels wave of coronavirus misinformation as users focus on popularity, not accuracy. <https://theconversation.com/social-media-fuels-wave-of-coronavirus-misinformation-as-users-focus-on-popularity-not-accuracy-135179>.
- Appiah, O., 2006. Rich media, poor media: the impact of Audio/Video vs. Text/Picture testimonial ads on browsers' evaluations of commercial web sites and online products. *J. Curr. Issues Res. Advert.* 28, 73–86. <https://doi.org/10.1080/10641734.2006.10505192>.
- Armstrong, P., 2023. As YouTube introduces revenue-sharing for shorts, the feature is rife with anti-LGBTQ vitriol, racism, misogyny, and COVID-19 misinformation — *mediamatters.org*. <https://www.mediamatters.org/google/youtube-introduces-revenue-sharing-shorts-feature-rife-anti-lgbtq-vitriol-racism-misogyny>.
- Asplund, J., Eslami, M., Sundaram, H., Sandvig, C., Karahalios, K., 2020. Auditing race and gender discrimination in online housing markets. *Proc. Int. AAAI Conf. Web Soc. Media* 14, 24–35. <https://doi.org/10.1609/icwsm.v14i1.7276>.
- Aubin Le Quéré, M., Schroeder, H., Randazzo, C., Gao, J., Epstein, Z., Perrault, S.T., Mimmo, D., Barkhuus, L., Li, H., 2024. LLMs as research tools: applications and evaluations in HCI data work. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7.
- Bakan, S., 2023. TikTok faces 'intensely difficult' battle against misinformation — *thenewdaily.com.au*. <https://www.thenewdaily.com.au/life/2023/10/16/tiktok-misinformation-moderation-eu>.
- Bandy, J., 2021. Problematic machine behavior: a systematic literature review of algorithm audits. *Proc. ACM Hum.-Comput. Interact.* 5, 1–34. <https://doi.org/10.1145/3449148>.
- bbc, 2012. The shining theories explored in spooky new documentary — *bbc.com*. <https://www.facebook.com/bbcnews>, <https://www.bbc.com/news/uk-scotland-highlands-islands-19789783>.
- Boeker, M., Urman, A., 2022. An empirical investigation of personalization factors on TikTok. In: *Proceedings of the ACM Web Conference 2022*, pp. 2298–2309.
- Boulianne, S., Lee, S., 2022. Conspiracy beliefs, misinformation, social media platforms, and protest participation. *Media Commun.* 10, <https://doi.org/10.17645/mac.v10i4.5667>.
- Brewster, J., Lorenzo Arvanitis, V., Pavilonis, Wang, M., 2022. Beware the 'New Google': TikTok's search engine pumps toxic misinformation to its young users. <https://www.newsguardtech.com/misinformation-monitor/september-2022/>.
- Brown, S., Davidovic, J., Hasan, A., 2021. The algorithm audit: scoring the algorithms that score us. *Big Data Soc.* 8, 205395172098386. <https://doi.org/10.1177/2053951720983865>.
- Bu, Y., Sheng, Q., Cao, J., Qi, P., Wang, D., Li, J., 2023. Combating online misinformation videos: characterization, detection, and future directions. In: *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 8770–8780.
- Caramanion, K.M., 2023. News verifiers showdown: a comparative performance evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in news fact-checking. *arXiv preprint arXiv:2306.17176*.
- Choi, E.C., Ferrara, E., 2023. Automated claim matching with large language models: empowering fact-checkers in the fight against misinformation. *arXiv preprint arXiv:2310.09223*.
- Clara Martiny, I.J., Cooper, L., 2022. Election misinformation thrives following social media platforms' shift to short-form video content — *isdglobal.org*. https://www.isdglobal.org/digital_dispatches/election-disinformation-thrives-following-social-media-platforms-shift-to-short-form-video-content/.
- Cuesta-Valiño, P., Gutiérrez-Rodríguez, P., Durán-Álamo, P., 2022. Why do people return to video platforms? Millennials and Centennials on TikTok. *Media Commun.* 10, <https://doi.org/10.17645/mac.v10i1.4737>.
- de Luna, E., 2022. Creators say YouTube shorts has a transphobia problem — *mashable.com*. <https://mashable.com/article/youtube-shorts-transphobic>.
- Dewan, A., 2022. Scientists SLAM joe rogan's podcast episode with Jordan peterson as 'absurd' and 'dangerous' | CNN — *edition.cnn.com*. <https://edition.cnn.com/2022/01/27/us/joe-rogan-jordan-peterson-climate-science-intl/index.html>.
- Ding, B., Qin, C., Liu, L., Bing, L., Joty, S., Li, B., 2022. Is GPT-3 a good data annotator? *arXiv preprint arXiv:2212.10450*.
- Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., Zhan, J., Zhang, H., 2011. Understanding the impact of video quality on user engagement. In: *Proceedings of the ACM SIGCOMM 2011 Conference, SIGCOMM '11*. Association for Computing Machinery, New York, NY, USA, pp. 362–373. <https://doi.org/10.1145/2018436.2018478>.
- Ecker, U.K.H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L.K., Brashier, N., Kendeou, P., Vraga, E.K., Amazeen, M.A., 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nat. Rev. Psychol.* 1, 13–29. <https://doi.org/10.1038/s44159-021-00006-y>.
- Elise Thomas, K.B., 2022. Algorithms as a weapon against women: how YouTube lures boys and young men into the 'Manosphere' — *isdglobal.org*. <https://www.isdglobal.org/isd-publications/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere/>.
- Enberg, J., 2022. US TikTok users by age, 2022. <https://www.emarketer.com/chart/256247/us-tiktok-users-by-age-2022-of-total>.
- Eriksson, M., Johansson, A., 2017. Tracking gendered streams. <https://doi.org/10.25595/1449>. <https://www.genderopen.de/handle/25595/1455>.
- Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., Moran, S., 2018. Falling for fake news: investigating the consumption of news via social media. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–10.
- Gabis, L.V., Attia, O.L., Goldman, M., Barak, N., Tefera, P., Shefer, S., Shaham, M., Lerman-Sagie, T., 2022. The myth of vaccination and autism spectrum. *Eur. J. Paediatr. Neurol.* 36, 151–158.
- Gilardi, F., Alizadeh, M., Kubli, M., 2023. ChatGPT outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.
- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., Wilson, C., 2013. Measuring personalization of web search. In: *Proceedings of the 22nd International Conference on World Wide Web*, pp. 527–538.
- Hannak, A., Soeller, G., Lazer, D., Mislove, A., Wilson, C., 2014. Measuring price discrimination and steering on e-commerce web sites. In: *Proceedings of the 2014 Conference on Internet Measurement*, pp. 305–318.
- Heseltine, M., Clemm von Hohenberg, B., 2024. Large language models as a substitute for human experts in annotating political text. *Res. Polit.* 11, 20531680241236239.
- Hoes, E., Altay, S., Bermeo, J., 2023a. Leveraging Chat-GPT for efficient fact-checking. *PsyArXiv*. April 3. vol. 3.
- Hoes, E., Altay, S., Bermeo, J., 2023b. Using ChatGPT to fight misinformation: ChatGPT nails 72% of 12,000 verified claims. *PsyArXiv*. April 3. vol. 3.
- Houli, D., Radford, M.L., Singh, V.K., 2021. "COVID19 is": the perpetuation of coronavirus conspiracy theories via Google autocomplete. *Proc. Assoc. Inf. Sci. Technol.* 58, 218–229. <https://doi.org/10.1002/pr2.450>.
- Hsu, T., 2022. Worries grow that TikTok is new home for manipulated video and photos. <https://www.nytimes.com/2022/11/04/technology/tiktok-deepfakes-disinformation.html>.

- Hu, D., Jiang, S., Robertson, R.E., Wilson, C., 2019. Auditing the partisanship of Google search snippets. In: *The World Wide Web Conference*, pp. 693–704.
- Hussein, E., Juneja, P., Mitra, T., 2020. Measuring misinformation in video search platforms: an audit study on YouTube. *Proc. ACM Hum.-Comput. Interact.* 4, 1–27. <https://doi.org/10.1145/3392854>
- Instagram, 2026. Help center — help.instagram.com. <https://help.instagram.com/173579827653028>.
- Iqbal, M., 2023. TikTok revenue and usage statistics (2023) — businessofapps.com. <https://www.businessofapps.com/data/tik-tok-statistics/>.
- Jane Lityvnenko, C.S., 2020. Here's a running list of false and misleading information about the election. <https://www.buzzfeednews.com/article/janelityvnenko/election-rumors-debunked>.
- Johnson, S.B., King, A.J., Warner, E.L., Aneja, S., Kann, B.H., Bylund, C.L., 2023. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr.* 7, pkad015.
- Juneja, P., 2023. Towards Understanding and Defending Against Algorithmically Curated Misinformation (Ph.D. thesis). University of Washington.
- Juneja, P., Bhuiyan, M.M., Mitra, T., 2023. Assessing enactment of content regulation policies: a post hoc crowd-sourced audit of election misinformation on YouTube. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–22.
- Keenan, C., 2022. An update on our work to counter misinformation — newsroom.tiktok.com. <https://newsroom.tiktok.com/en-au/an-update-on-our-work-to-counter-misinformation-au>.
- Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., Mislove, A., 2015. Location, location, location: the impact of geolocation on web search personalization. In: *Proceedings of the 2015 Internet Measurement Conference*, pp. 121–127.
- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M.B., Ghosh, S., Gummadi, K.P., Karahalios, K., 2017. Quantifying search bias: investigating sources of bias for political searches in social media. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 417–432.
- Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., Shafiq, Z., 2019. Measuring political personalization of Google news search. In: *The World Wide Web Conference*, pp. 2957–2963.
- Liedke, J., Wang, L., 2023. Social media and news fact sheet. <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>.
- Lim, C., 2018. Checking how fact-checkers check. *Res. Polit.* 5, 2053168018786848.
- Lundy, M., 2023. New study reveals TikTok's spread of COVID-19 misinformation. <https://ischool.illinois.edu/news-events/news/2023/06/new-study-reveals-tiktoks-spread-covid-19-misinformation>.
- Lityvnenko, J., 2020. Here's a running list of the latest hoaxes spreading about the coronavirus — buzzfeednews.com. <https://www.buzzfeednews.com/article/janelityvnenko/coronavirus-fake-news-disinformation-rumors-hoaxes>.
- Madraki, G., Grasso, I., Ojala, J.M., Liu, Y., Matthews, J., 2021. Characterizing and comparing COVID-19 misinformation across languages, countries and platforms. In: *Companion Proceedings of the Web Conference 2021*, pp. 213–223.
- McCrosky, J., Geurkink, B., 2021. YouTube regrets: a crowdsourced investigation into YouTube's recommendation algorithm. Mozilla Found. Retrieved November 15, 2021.
- McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282.
- Menat, T., 2022. TikTok removes more than 102mln videos for violating its community guidelines in q1 2022 — zawya.com. <https://www.zawya.com/en/press-release/companies-news/tiktok-removes-more-than-102mln-videos-for-violating-its-community-guidelines-in-q1-2022-ty418zrl>.
- Metaxa, D., Park, J.S., Landay, J.A., Hancock, J., 2019. Search media and elections. *Proc. ACM Hum.-Comput. Interact.* 3, 1–17. <https://doi.org/10.1145/3359231>
- Micallef, N., Sandoval-Castañeda, M., Cohen, A., Ahamad, M., Kumar, S., Memon, N., 2022. Cross-platform multimodal misinformation: taxonomy, characteristics and detection for textual posts and videos. *Proc. Int. AAAI Conf. Web Soc. Media* 16, 651–662. <https://doi.org/10.1609/icwsm.v16i1.19323>
- Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N., 2012. Detecting price and search discrimination on the internet. In: *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pp. 79–84.
- Mohan, N., 2021. Perspective: tackling misinformation on YouTube — blog.youtube. <https://blog.youtube/inside-youtube/tackling-misinfo/> (Accessed 22 June 2024).
- Morrow, A., 2022. Joe rogan's path from reality host to covid controversy king | CNN business — edition.cnn.com. <https://edition.cnn.com/2022/02/01/media/joe-rogan-spotify-controversy/index.html>.
- Newton, C., 2017. How YouTube perfected the feed. <https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-recommendation-personalized-feed>.
- O'Connor, C., 2021. How TikTok sounds are used to fuel anti-vaccine fears. <https://www.isdglobal.org/digital-dispatches/how-tiktok-sounds-are-used-to-fuel-anti-vaccine-fears> (Accessed 6 January 2024).
- Olteanu, A., Diaz, F., Kazai, G., 2020. When are search completion suggestions problematic? *Proc. ACM Hum.-Comput. Interact.* 4, 1–25. <https://doi.org/10.1145/3415242>
- OpenAI, 2023. GPT-4 system card. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>.
- Organisation, W.H., 2020. COVID-19 vaccines and vaccination explained — covid19info-vaccines.com. <https://www.covid19infovaccines.com/>.
- O'Sullivan, N.J., Nason, G., Manecksha, R.P., O'Kelly, F., 2022. The unintentional spread of misinformation on 'TikTok': a paediatric urological perspective. *J. Pediatr. Urol.* 18, 371–375. <https://doi.org/10.1016/j.jpuro.2022.03.001>
- Papadamous, K., Zannettou, S., Blackburn, J., De Cristofaro, E., Stringhini, G., Sirivianos, M., 2020. "It is just a flu": assessing the effect of watch history on YouTube's pseudoscientific video recommendations. [arxiv:2010.11638](https://arxiv.org/abs/2010.11638).
- Park, J., Wisniewski, P., Singh, V., 2024. Leveraging large language models (LLMs) to support collaborative Human-AI online risk data annotation. [arXiv preprint arXiv:2404.07926](https://arxiv.org/abs/2404.07926).
- Pasquale, F., 2010. Beyond innovation and competition: the need for qualified transparency in internet intermediaries. *Northwest. Univ. Law Rev.* 104, <https://doi.org/10.2139/ssrn.1686043>
- Pasquale, F.A., 2011. Restoring transparency to automated authority. *J. Telecomm. High Technol. Law* 9, 235.
- Paul, K., 2022. 'We risk another crisis': TikTok in danger of being major vector of election misinformation. <https://www.theguardian.com/technology/2022/oct/24/tiktok-election-misinformation-voting-politics>.
- Qi, P., Bu, Y., Cao, J., Ji, W., Shui, R., Xiao, J., Wang, D., Chua, T.-S., 2022. FakeSV: a multimodal benchmark with rich social context for fake news detection on short video platforms. [arxiv:2211.10973](https://arxiv.org/abs/2211.10973).
- Qiu, F., Cui, Y., 2011. A quantitative study of user satisfaction in online video streaming. In: 2011 IEEE Consumer Communications and Networking Conference (CCNC), pp. 410–414. <https://doi.org/10.1109/CCNC.2011.5766502>
- Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjeh, R., Robertson, C.E., Van Bavel, J.J., 2024. GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl. Acad. Sci.* 121 (34), e2308950121.
- Rogers, R., 2021. Marginalizing the mainstream: how social media privilege political information. *Front. Big Data* 4, <https://doi.org/10.3389/fdata.2021.689036>
- Roy, S., Ayalon, L., 2019. Age and gender stereotypes reflected in Google's "Autocomplete" function: the portrayal and possible spread of societal stereotypes. *Gerontologist* 60, 1020–1028. <https://doi.org/10.1093/geront/gnz172>
- Sallam, M., Dababseh, D., Yaseen, A., Al-Haidar, A., Ababneh, N.A., Bakri, F.G., Mahafzah, A., 2020. Conspiracy beliefs are associated with lower knowledge and higher anxiety levels regarding COVID-19 among students at the university of Jordan. *Int. J. Environ. Res. Public Health* 17, 4915. <https://doi.org/10.3390/ijerph17144915>
- Sandvig, C., Hamilton, K., Karahalios, K., Langbort, C., 2014. Auditing algorithms: research methods for detecting discrimination on internet platforms. *Data Discrim.: Converting Crit. Concerns Product. Inq.* 22, 4349–4357.
- Sardarizadeh, S., 2019. Instagram fact-check: can a new flagging tool stop fake news? — bbc.com. <https://www.bbc.com/news/blogs-trending-49449005>.
- Sillito, D., 2022. Teens shun traditional news channels for TikTok and instagram, ofcom says — bbc.com. <https://www.bbc.com/news/entertainment-arts-62238307>.
- Singh, H., 2022. Darning report reveals depths of TikTok misinformation — the-newdaily.com.au. <https://www.thenewdaily.com.au/life/tech/2022/09/25/tiktok-misinformation-report>.
- Southwell, B.G., Brennen, J.S.B., Paquin, R., Boudewyns, V., Zeng, J., 2022. Defining and measuring scientific misinformation. *Ann. Am. Acad. Pol. Soc. Sci.* 700, 98–111.
- Spezzano, F., Shrestha, A., Fails, J.A., Stone, B.W., 2021. That's fake news! Reliability of news when provided title, image, source bias & full article. *Proc. ACM Hum.-Comput. Interact.* 5, 1–19. <https://doi.org/10.1145/3449183>
- Spinelli, L., Crovella, M., 2020. How YouTube leads privacy-seeking users away from reliable information. In: *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pp. 244–251.
- Srba, I., Moro, R., Tomlein, M., Pecher, B., Simko, J., Stefancova, E., Kompan, M., Hrkova, A., Podrouzek, J., Gavornik, A., Bielikova, M., 2023. Auditing YouTube's recommendation algorithm for misinformation filter bubbles. *ACM Trans. Recomm. Syst.* 1, 1–33. <https://doi.org/10.1145/3568392>
- Sundar, S.S., 2000. Multimedia effects on processing and perception of online news: a study of picture, audio, and video downloads. *Journal. Mass Commun. Q.* 77, 480–499. <https://doi.org/10.1177/107769900007700302>
- Sundar, S.S., Molina, M.D., Cho, E., 2021. Seeing is believing: is video modality more powerful in spreading fake news via online messaging apps? *J. Comput.-Mediat. Commun.* 26, 301–319. <https://doi.org/10.1093/jcmc/zmab010>
- TikTok, 2021. Terms of service | TikTok — tiktok.com. <https://www.tiktok.com/legal/page/row/terms-of-service/en#terms-eea>.
- Tomasik, E., Matsa, E.M., 2025. 1 in 5 Americans now regularly get news on TikTok, up sharply from 2020. *Pew Res. Cent.* <https://www.pewresearch.org/short-reads/2025/09/25/1-in-5-americans-now-regularly-get-news-on-tiktok-up-sharply-from-2020/>.
- Tomlein, M., Pecher, B., Simko, J., Srba, I., Moro, R., Stefancova, E., Kompan, M., Hrkova, A., Podrouzek, J., Bielikova, M., 2021. An audit of misinformation filter bubbles on YouTube: bubble bursting and recent behavior changes. In: *Proceedings of the 15th ACM Conference on Recommender Systems*, pp. 1–11.
- Törnberg, P., 2023. ChatGPT-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. [arxiv:2304.06588](https://arxiv.org/abs/2304.06588).
- Trielli, D., Diakopoulos, N., 2019. Search as news curator: the role of Google in shaping attention to news information. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.
- Tseng, Y.-C., Bi, N., Chang, Y.-J., Yuan, C.W., 2022. Investigating correction effects of different modalities for misinformation about COVID-19. In: *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 54–58.
- Tucker, E., 2022. TikTok's search engine repeatedly delivers misinformation to its majority-young user base, report says. <https://edition.cnn.com/2022/09/18/business/tiktok-search-engine-misinformation/index.html>.
- van Kampen, K., Laski, J., Herman, G., Chan, T.M., 2022. Investigating COVID-19 vaccine communication and misinformation on TikTok: cross-sectional study. *JMIR Infodemiol.* 2, e38316. <https://doi.org/10.2196/38316>
- Venkatagiri, S., Schafer, J.S., Prochaska, S., 2023. The challenges of studying misinformation on video-sharing platforms during crises and mass-convergence events. [arxiv:2303.14309](https://arxiv.org/abs/2303.14309).

- Wang, X., Kim, H., Rahman, S., Mitra, K., Miao, Z., 2024. Human-LLM collaborative annotation through effective verification of LLM labels. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–21.
- Weimann, G., Masri, N., 2020. Research note: spreading hate on TikTok. *Stud. Confl. Terror.* 46, 752–765. <https://doi.org/10.1080/1057610X.2020.1780027>
- Wittenberg, C., Tappin, B.M., Berinsky, A.J., Rand, D.G., 2021. The (minimal) persuasive advantage of political video over text. *Proc. Natl. Acad. Sci.* 118, <https://doi.org/10.1073/pnas.2114388118>
- Wong, Q., 2022. TikTok is a misinformation Minefield. Don't get tripped up. <https://www.cnet.com/news/misinformation/tiktok-is-a-misinformation-minefield-dont-get-tripped-up/>.
- Xie, X., Du, Y., Bai, Q., 2022. Why do people resist algorithms? From the perspective of short video usage motivations. *Front. Psychol.* 13, <https://doi.org/10.3389/fpsyg.2022.941640>
- Yadav, A., Phillips, M.M., Lundeberg, M.A., Koehler, M.J., Hilden, K., Dirkin, K.H., 2011. If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *J. Comput. High. Educ.* 23, 15–37. <https://doi.org/10.1007/s12528-011-9042-y>
- Yang, S., Zhao, Y., Ma, Y., 2019. Analysis of the reasons and development of short video application-taking TIK tok as an example. In: *Proceedings of the 2019 9th International Conference on Information and Social Science (ICISS 2019)*, Manila, Philippines, pp. 12–14.
- Yeung, A., Ng, E., Abi-Jaoude, E., 2022. TikTok and Attention-Deficit/Hyperactivity disorder: a cross-sectional study of social media content quality. *Can. J. Psychiatry* 67, 899–906. <https://doi.org/10.1177/07067437221082854>
- YouTube May Have Misinformation Blind Spots, 2022. Researchers say (published 2022) — nytimes.com. <https://www.nytimes.com/2022/11/05/technology/youtube-misinformation.html>.
- Zhao, Z., 2021. Analysis on the “Douyin (TikTok) mania” phenomenon based on recommendation algorithms. *E3S Web Conf.* 235, 03029. <https://doi.org/10.1051/e3sconf/202123503029>