Amsterdam University Press

COMPUTATIONAL COMMUNICATION RESEARCH 8.4 (2026) 1–29
HTTPS://DOI.ORG/10.5117/CCR2026.4.2.LI

Stitching, dueting, and playing with science on TikTok: An AI-powered multimodal approach to understanding interactive science videos and audience engagement

Yuhan Li[a], Annie Li Zhang[b], Hang Lu[a]

[a] *Department of Communication and Media, University of Michigan, Ann Arbor, MI, USA*

[b] *Center for an Informed Public, University of Washington, Seattle, WA, USA*

Abstract

TikTok offers interactive features, such as stitch, duet, and reply-to, that enable creators to respond to others and co-create science content, but it remains unclear whether these features translate into participatory science communication in practice. We developed an AI-powered framework to analyze 2,796 TikTok science short videos with 31,782 image frames and 2,796 audio files, examining how interactive features intersect with communicator identity, communication objectives, and audio-visual presentation strategies. We found that interactive videos were a small share of the overall science videos, yet they showed both promise and risks of participatory science communication: they were actively adopted by scientists and influencers and often used to inform the public and debunk misinformation, but they also facilitated the circulation of pseudoscience. Using latent class analysis, we identified four recurring styles of interactive videos: professional-led explanatory science, ordinary user-led personal commentary on science, influencer-led entertainment or experimental science, and ordinary user-led visual science. This study enriches our understanding of participatory science communication on TikTok and offers an AI-powered analysis framework for studying multimodal science messages on short-video platforms.

Keywords: short-video, science communication, TikTok, multimodal communication, public engagement

Introduction

"What you saw in that video was a lady taking her little kid's shoes, putting water into it and then putting it into the freezer, which is incredibly smart, because what's gonna happen is that shoes gonna get bigger. And I'm gonna

Manuscript received 2025-12-01, accepted 2026-04-13, published 2026-05-17

explain to you why.” In this video, a TikTok influencer with over 600,000 followers stitches together a video published by another user to explain the physics behind the phenomenon, receiving over 4,000 likes.

This example highlights a growing trend of science communication on TikTok that goes beyond standalone short videos. Through platform-embedded interactive features such as *stitch*, *duet*, and *reply-to*, creators can co-create content, remix ideas, and engage with other videos, supporting a wide array of science content such as knowledge sharing (Nguyen & Diederich, 2023), misinformation debunking, and science advocacy (Hautea et al., 2021). Meanwhile, creators actively use TikTok’s editing capabilities, such as memes, filters, and music, to create multimodal science content. Unlike the traditional one-way dissemination paradigm, these interactive features enable a more engagement-oriented paradigm that emphasizes interaction between multiple actors (Zimmerman et al., 2025).

Indeed, with its unique algorithmic feeds and audio-visual culture, short-video platforms such as TikTok have rapidly become central arenas where younger generations encounter and participate in science (Zeng et al., 2021). However, the interactive and multimodal nature of science short videos remains theoretically and empirically underexplored. Drawing on a dataset of 2,796 science-related videos encompassing image frames, audio files, texts, and engagement data, we compare videos using stitch/duet/reply-to features with a randomly sampled set of non-interactive videos to examine (1) which groups of communicators are more likely to adopt interactive features, (2) which communication objectives are more common in interactive videos, (3) what audio-visual presentation strategies can characterize interactive versus non-interactive videos, and (4) how these features jointly shape audience engagement. Using a comparative perspective, this study aims to advance our understanding of participatory science communication on TikTok and offer methodological insights for leveraging multimodal AI models to analyze short videos.

Literature Review

Evolving models of science communication: from deficit to dialogue and to participation

Science communication has evolved through several conceptual models over the past decades (Bucchi, 2008). The traditional conception of science communication is the deficit model, which posits a one-way transmission from experts to uninformed publics (Lewenstein, 2003). However, critics noted that this approach was paternalistic and failed to account for public values and local knowledge. In response, scholars advocated for the dialogue model, emphasizing two-way communication and seeking public input on science-related issues (Bucchi, 2008). Nonetheless, the dialogue model was criticized for remaining expert-controlled, with elites and scientists setting agendas (Trench, 2008).

More recently, a public participation model has emerged, highlighting the role of publics in scientific knowledge co-production. Unlike the traditional deficit or dialogue models, the participation model is "multidirectional, open-ended, and potentially subject to conflict" (Bucchi, 2008, p. 70). Participatory science communication happens "when scientists and/or science communicators interact with various publics in a dynamic process where different forms of knowledge and experiences are acknowledged, shared, valued and negotiated, and where power relations are levelled." (Metcalfe et al., 2022, p. 5)

Digital platforms have accelerated these shifts by lowering participation barriers (Zeng et al., 2021). TikTok's algorithms disseminate science-related videos not only to viewers with existing interests in science but also to broader audiences who may not actively seek such content (Nguyen & Diederich, 2023; Rein, 2023). With affordances such as high editability, visibility, and personalization, short-video platforms are reshaping science communication from passive knowledge dissemination towards co-production. Within this context, our study focuses on three interactive features prevalent on TikTok that remain underexplored in science communication research: stitch, duet, and reply-to.

We define *interactive science videos* as those employing any of the three features. Stitch enables users to add up to a five-second clip from another video to the beginning of their own; duet places two videos side-by-side in a split-screen format; reply-to allows creators to respond to comments with a new video. Theoretically, these features could achieve participatory ideals by facilitating interaction and idea exchange among multiple actors (Nguyen & Diederich, 2023) and enabling every user to respond to, recontextualize, or challenge existing science content. However, a critical question remains: Do these features genuinely enable participatory science communication across diverse actors? Or do they simply reproduce deficit and dialogue models? We therefore examine whether and to what extent TikTok's interactive features allow for participatory science communication in practice. Given limited knowledge of these features in science communication contexts, we begin with an exploratory question assessing their prevalence on TikTok:

RQ1. How prevalent are interactive features (i.e., stitch, duet, and reply-to) used in science-related videos on TikTok?

Communicative features and multimodal configurations of interactive science videos

Interactive science videos on TikTok present a distinct form of multimodal communication involving visual, auditory, and textual elements (Bainotti & Peeters, 2025; Divon & Eriksson Krutrök, 2024). TikTok's editability enables users to display and visualize scientific ideas through multiple layers of visuals, sounds, and filters, fostering memetic and creative construction of science (Zeng et al., 2021). Its easy-to-use interfaces introduce distinct aesthetic styles and allow diverse actors to participate in public science communication, expanding beyond the traditional expert-driven dissemination mode and fostering creative expressions surrounding science (Rein, 2023; Zimmerman et al., 2025).

Existing literature has largely focused on memetic science events and their aesthetic characteristics (Chen et al., 2022; Zeng et al., 2021), with less attention to how interactive features afford participatory science communication. Drawing upon a large-scale multimodal dataset of TikTok science-related videos, we develop a comprehensive analysis framework encompassing the communicative features (e.g., who creates the content and what they aim to achieve) and audio-visual presentation strategies of videos, introducing a typology that recognizes how meaning is constructed through visual, audio, and textual elements (Kress, 2009).

Communicator identity

Traditional deficit and dialogue models position scientists, journalists, and institutions as primary science communicators (Lewenstein, 2003), while the participation model highlights diverse actors' roles in knowledge co-production (Bucchi, 2008, p. 69). Prior research shows that influencers and ordinary users frequently leverage duets or stitches to connect, collaborate, and create new content to enhance their audience reach (Kaye, 2022). While social media platforms lower barriers to science participation, we know little about the extent to which lay participants use interactive features to engage in scientific content production.

Meanwhile, scientists, journalists, and institutions continue adapting their roles in response to the evolving digital media environment (Fahy & Nisbet, 2011) and increasingly leverage popular culture for legitimization

(Zeng et al., 2021). Given the potential of interactive features in engaging audiences, they may offer scientists and scientific institutions unique opportunities to present science in ways that align with TikTok's entertainment-driven culture. However, it remains unclear to what extent scientists adopt these features as influencers or everyday users do. Knowing who adopts interactive features in science communication would shed light on whether digital technologies enable genuine participatory science communication or simply provide new channels for expert-dominant science discussions. We therefore propose:

RQ2. Which groups of communicators are more likely to adopt interactive features?

Communication objectives

Interactive features can enable communication objectives reflecting different models of science communication (Evans et al., 2017). Traditional deficit and dialogue models prioritize objectives such as informing the public about science and defending science from misinformation (Dudo & Besley, 2016, p. 8), positioning scientists as authorities mitigating public deficits. By contrast, the participation model suggests a more relational orientation, where "different forms of knowledge and experiences are acknowledged, shared, valued and negotiated" by scientists and publics (Metcalfe et al., 2022, p. 5). Rather than simply transmitting information, relational objectives highlight building connections with audiences and fostering mutual engagement. We thus anticipate that the participation model would involve new objectives beyond merely informing the public.

Recent TikTok research shows that creators often use interactive features for varied purposes: informing, entertaining, challenging viewpoints, and producing humor (Divon & Eriksson Krutrök, 2024). Zeng et al. (2021) showed that science videos actively present science as entertaining activities, motivating users to recreate the content. Moreover, instead of just passively receiving opinions from experts, TikTok users also contribute subjective viewpoints or lived experiences to science-related discussions. However, the multidirectional, open-ended nature of participatory communication also introduces risks. Promoting pseudoscience or conspiracy exemplifies the "conflict" inherent in participation (Bucchi, 2008, p. 70): lowered barriers enable not only diverse perspectives but also content challenging scientific consensus (Scheufele & Krause, 2019).

The flexibility of interactive features allows science communication to shift from one-way transmission to multidirectional processes involving multiple communicators pursuing varied objectives (Trench, 2008). If interactive videos exhibit diverse objectives that go beyond informational goals, it suggests movement toward participatory science communication. Accordingly, we ask:

RQ3. Compared with videos without interactive features, which communication objectives are more likely to be expressed in interactive science videos?

Audio-visual presentation strategies

Traditional science videos rely on didactic formats, such as formal lectures and graphics, to reinforce authority and establish professional images. While we have limited knowledge about what participatory science communication would look like, we expect its presentation would become more diverse and individualized, as it involves scientists and various publics to share, exchange, and negotiate scientific knowledge (Metcalfe et al., 2022). TikTok encourages visual layering, split screens, embedded captions, voice overlays, and synchronized music, all of which function as semiotic resources shaping video perception (Kress, 2009). With the advantage of TikTok's high editability, creators utilize built-in visual and audio features to create diverse content with different configurations. Prior studies suggest that audio-visual presentations of short videos impact audiences' psychological experiences of immersion (Wang, 2020). It could be expected that creators may use particular presentation styles to attract audiences' attention or build personalized branding when dueting or stitching TikTok videos.

Among various visual presentation strategies, our study examines three strategies particularly relevant to science communication: first-person host, visual aids, and live experiments. First, the "point-of-view" (POV) is an aesthetic format that often goes viral on TikTok, where videos are filmed from a first-person perspective and make audiences feel as if they are the main character in a scenario (Divon & Eriksson Krutrök, 2024). With the first-person perspective, people might consider the storytelling more trustworthy and feel more immersed (Wang, 2020). By highlighting the role of the video host, a first-person host could increase the personalization of science communication (Zeng et al., 2021).

Visual aids have become an important means for facilitating the understanding of complex scientific concepts, especially for the lay public (Rodríguez Estrada & Davis, 2015). While other video-sharing platforms also enable visuals, TikTok's high editability and templates incentivize creators to apply visual aids more actively. When interacting with audiences, appropriate visual aids might increase perceived quality, enhance processing fluency, and help audiences better understand arguments presented by video hosts.

Hands-on experimentation is one of the most effective ways to learn, engage with, and retain scientific information (Jones & Stapleton, 2017). TikTok enables live demonstrations that transcend geographical limitations between scientists and audiences. Stitch or duet videos might use live experiments to respond to commenters' questions or requests, a strategy shown to strengthen audience engagement (Habibi & Salim, 2021).

The audio presentation sits at the center of stitch and duet videos on TikTok. Besides speech-dominated videos that only include monologues or conversations, "audio memes" have long been popularized on TikTok and present an "aural turn" in meme ecologies (Abidin & Kaye, 2021). TikTokers often rely on the lyrics of specific songs to tell a story and engage with other original audio templates through duets and replies (Abidin & Kaye, 2021; Kaye, 2022). More importantly, music videos are seen as helpful tools for science communication because of their short format, approachability, and shareability (Rein, 2023). TikTok also allows creators to add background music to mainly speech videos. Considering the complexity of audio-visual presentation strategies, we ask the following research question:

RQ4. Compared with videos without interactive features, which (a) visual and (b) audio presentation strategies are more likely to appear in interactive science videos?

Co-occurring patterns of communicative and audio-visual presentation strategies

Prior research on short videos has often focused on either textual-only or visual-only data (Huang et al., 2010). Yet communicator identity, communication objective, and visual- and audio-level features "ultimately feed into the same kind of construction process wherein a coherent interpretation is formed" (Geise & Baden, 2015, p. 47). Similarly, multimodal theory conceptualizes communication as the orchestration of visual, auditory, and textual modes in meaning-making processes (Kress, 2009). Indeed, recent studies on short-video platforms are increasingly combining audio-visual data to derive a parsimonious typology of short-video production (Lu & Shen, 2023). Integrating advances from visual communication, we aim to move beyond variable-level analysis to achieve a holistic understanding of interactive science videos. Thus, we propose:

RQ5. Which communicative features and audio-visual presentation strategies tend to be employed together in interactive science videos?

Interactive science communication and audience engagement

Different communicative features and audio-visual presentation strategies of short videos shape information perception and are associated with varying levels of audience responses (Geise & Baden, 2015). For example, communicator identity plays an important role in this process: Zhang and Lu (2023) found that people rated scientists on X as more authentic and qualified than science influencers, while Zeng et al. (2021) reported that science institutional accounts on TikTok produced more engaging content than other communicators. Although Besley and Dudo (2022) identified communication objectives that facilitate public engagement, how these objectives perform in practice remains underexplored. While existing studies indicate experts' relative advantages in engaging the audience and outline common communication objectives, systematic evidence linking communicator identity, communication objectives, and audience engagement remains limited, particularly on short-video platforms like TikTok, where traditional power dynamics between science and the public have shifted.

Multimodal presentation strategies also shape audience engagement outcomes. Cheng and Li (2024) found that on TikTok, a character talking or dancing directly to the viewer in front of the camera can amplify audiences' psychological experiences of immersion, leading to higher levels of engagement. Habibi and Salim (2021) found that, compared to the lecture-style videos on TikTok, videos showing experimentation received significantly higher engagement. Regarding audio presentation, Crowther et al. (2016) showed that music versions of science videos were perceived as more enjoyable and were more likely to be revisited and shared than non-musical videos. These findings suggest that both communicative and audio-visual presentation factors shape audience engagement with science. Building on the above, we ask:

RQ6. Which communicative features and audio-visual presentation strategies are associated with audience engagement in interactive science videos?

Overall, communicator identity, communication objectives, and audio-visual presentation strategies serve as complementary indicators of whether TikTok's interactive features are used in ways that tap into participatory science communication. If interactive science videos are created by a broader range of communicators beyond traditional experts and institutions, pursue objectives beyond informing the public, and employ more individualized, diverse, and accessible presentation styles, this would suggest movement toward a more multidirectional, open-ended, and participatory model.

Framed this way, this study sheds light on how these communicative features and audio-visual presentation strategies are orchestrated together and associated with audience engagement, and examines the potential of TikTok's interactive features in expanding public participation in science communication.

Method

Data collection and preprocessing

The entire workflow of this study is described in Figure 1. We first used the TikTok Research API to search videos with the hashtag #science that were published in the United States between January 1, 2024, and December 31, 2024. This search resulted in 115,278 videos. Because videos tagged with #science do not necessarily contain scientific content, we filtered the dataset using a BERT-based classifier fine-tuned to distinguish science-related videos from non-science content (Precision = 0.83, Recall = 0.82, F1 = 0.83). The model was implemented with BertForSequenceClassification (Transformers v4.42.4) and trained on our labeled training dataset (see Supplementary Materials A for training details). Video descriptions and transcripts were used as input for classification, resulting in 34,348 science-related videos.

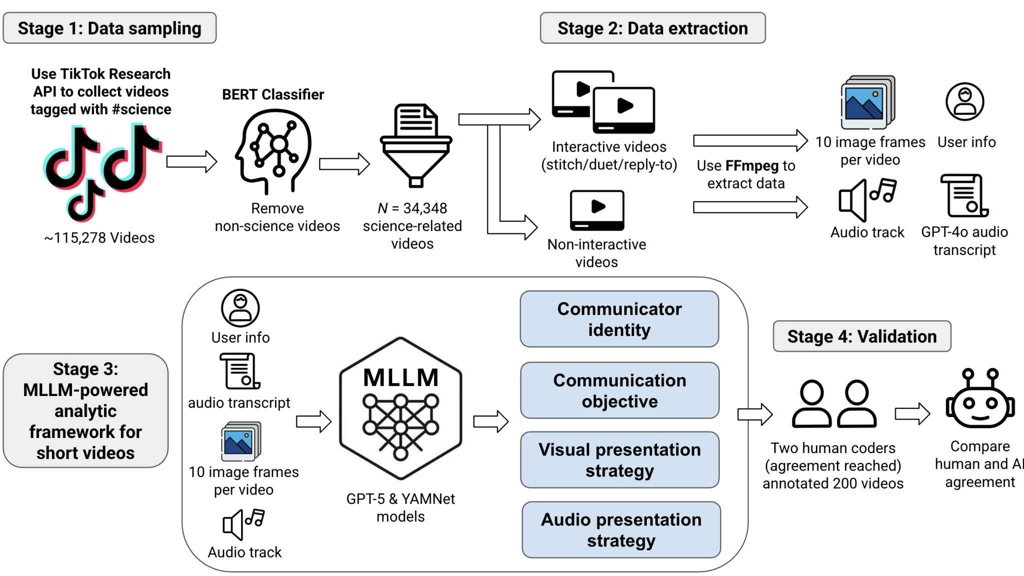


Figure 1: The workflow of this study.

Identifying interactive and non-interactive science videos

We operationalized interactive science communication as videos that used TikTok's "stitch," "duet," or "reply-to" functions. When creators use these features, TikTok automatically adds standardized attribution text (e.g., "#stitch with @xxx", "#duet with @xxx", "Replying to @xxx") to these videos, which cannot be removed from the video description. Thus, we identified interactive science videos using a dictionary-based approach: videos containing any of the above attribution texts in their video descriptions are classified as interactive science videos, and the specific interactive strategy was coded. This procedure yielded 1,538 interactive science videos. We then retrieved user-level metadata (e.g., verified status and follower count) via the TikTok Research API. Since some accounts were not publicly available, 1,509 interactive videos remained after data preprocessing. To construct a comparison group, we randomly sampled 2,000 non-interactive videos from the remaining corpus. We used simple random sampling, implemented in R with a fixed random seed to ensure reproducibility. After data preprocessing and account availability checks, 1,801 non-interactive videos with user-level metadata remained.

Collecting the image frames, audio tracks, and transcripts

To conduct multimodal analysis, we downloaded all sampled videos using a Python-based web scraping script. We used FFprobe (FFmpeg multimedia framework) to extract video duration (in seconds) from metadata (Tomar, 2006). Videos with invalid durations were recorded as errors and excluded from subsequent analysis. Using FFmpeg, we extracted both image frames and audio tracks from each video. To capture representative visual content across the entire clip, we sampled 10 image frames per video by dividing the video's duration into 10 equal segments and capturing one frame at each timestamp. For extremely short clips, we imposed a minimum safe duration of 0.50 seconds to avoid extracting out-of-range frames. In parallel, we converted audio tracks into audiowave files for analysis.

In total, we collected 1,319 audiowaves and 14,821 image frames for interactive videos and 1,477 audiowaves and 16,961 image frames for non-interactive videos. To obtain audio transcripts, we processed the audiowaves using GPT-4o Transcribe, a state-of-the-art speech-to-text model. We used the langdetect Python package to detect the language of video transcripts. The results confirmed that 2,624 out of 2,796 videos were in English, with small amounts of non-English content. We saved the transcripts returned by GPT-4o and retained the 1,319 interactive science videos and 1,477 non-interactive science videos for the following analyses.

MLLM-powered Analysis Framework for Short Videos

We used GPT-5, a state-of-the-art multimodal large language model (MLLM), to annotate communicative features and audio-visual presentation strategies because these variables are multimodal and vary widely across creators and videos. It is difficult to accurately and efficiently capture these variables at scale with traditional supervised machine learning models. Because of its strong performance in multimodal understanding across various contexts, MLLM becomes the most suitable approach for our study purposes (OpenAI, 2025). While LLM-based text classification is increasingly common in communication research (Stoll et al., 2025), applying MLLMs to short-video annotation remains less established and requires careful validation. Given the above, we developed an AI-powered multimodal framework with sampling, annotation, and validation procedures, described below.

Independent variables

Communicator identity was classified based on publicly available profile information (the number of followers and user bio description) compiled for each unique creator. For privacy concerns, we removed the user handle from later analysis and recoded follower count as a binary indicator (below 10,000 vs. 10,000 or above). GPT-5 received only de-identified inputs necessary for classifying communicator identity. Our coding scheme distinguishes communicators along two dimensions. First, we differentiate expert and authoritative voices such as scientists and journalists from non-experts, which has long been a traditional distinction in science communication research (Hautea et al., 2021). Second, we use the 10,000-follower threshold to distinguish established content creators from ordinary users. Creators with over 10,000 followers typically have sustained audiences and personal influence (TikTok, 2026). We provided user data to GPT-5 and prompted it to categorize each creator as: 1) news media/journalist: individual journalists or professional news outlets, 2) scientist/expert: individuals or institutions that show scientific expertise or education, 3) influencer/celebrity: non-expert users with more than 10,000 followers; and 4) ordinary user: non-expert users with fewer than 10,000 followers.

We identified the primary *communication objective* expressed in video content through an iterative, model-assisted coding process. Drawing on

Dudo and Besley's (2016) framework of science communication objectives (informing the public and defending science from misinformation) and Zeng et al.'s (2021) analysis of TikTok's entertainment-driven science communication, we initially prompted GPT-5 to label 200 randomly sampled videos using four categories: 1) informing the public about science, 2) defending science from misinformation, 3) entertaining or performing with science, and 4) open-coding. Through three rounds of iterative coding, two additional recurring objectives emerged: 5) promoting pseudoscience or conspiracy theories (i.e., presenting scientifically unfounded claims) and 6) sharing personal opinions or commentary about science (i.e., expressing subjective viewpoints or lived experiences). An "other" category was retained for videos not clearly fitting any classification. Based on the video description and transcript, GPT-5 classified all videos using the final scheme.

Visual presentation strategies were identified based on the extracted image frames (see Figure 2). Drawing on prior research on digital science communication and TikTok (Divon & Eriksson Krutrök, 2024; Zeng et al., 2021), we developed a coding scheme capturing common TikTok visual styles and used GPT-5 to annotate each video. We provided the sampled frames of each video to GPT-5 and prompted it to classify videos into the following categories: 1) first-person host: a person faces the camera and talks or presents directly to the audience; 2) visual aids: slides, infographics, diagrams, or research materials conveying information; 3) live experiment: visible experimentation, object manipulation, or process illustration; and 4) other: any alternative visual presentation styles. Since TikTok videos often apply more than one visual framing, GPT-5 was allowed to assign multiple labels per video.

Audio presentation strategies were identified using Google's YAMNet audio classification model, which estimates the presence of everyday sounds, including human speech and music (Hershey et al., 2017). We aggregated YAMNet's outputs into two broader dimensions: speech-related and music-related, producing a speech score and music score ranging from 0 to 1 for each video. We then categorized videos into four mutually exclusive groups based on the following thresholds: speech-dominant videos were those with a speech score > 0.50 and higher than the music score; music-dominant videos had a music score > 0.50 and higher than the speech score. We used 0.50 as a high-confidence cutoff to identify cases where one audio source clearly dominates. Mixed music and speech captures a common TikTok format in which creators speak clearly over low-volume background music. Accordingly, we required speech > 0.50 but used a lower music threshold

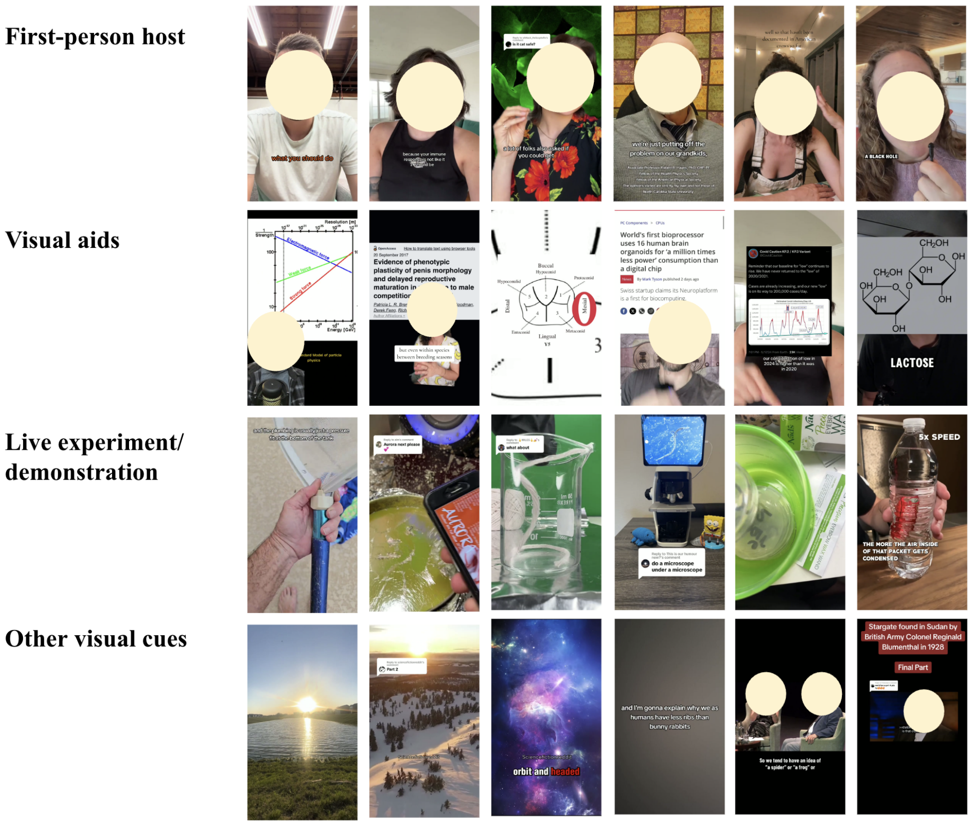


Figure 2: Examples of visual presentation strategies.

to detect low-volume music that YAMNet often assigns lower probabilities. Videos that did not meet any criterion were labeled as uncertain (see Supplementary Materials D for details).

We validated the AI-powered classification results as follows: two trained coders first discussed the codebook together and independently coded a random sample of 30 videos from the full corpus. Discrepancies in their coding were discussed and resolved until acceptable agreement was achieved. The coders then coded a different random sample of 200 videos (100 videos per coder). We reported the inter-coder reliability between the BERT-, MLLM-, or YAMNet-generated labels and human-coded results in Table 1.

	Intercoder reliability between human coders	Reliability between human- and machine-coded results (Krippendorff's Alpha)
Communicator identity	0.840	0.903
Communication objective	0.745	0.733
Visual presentation strategy		
First-person host	1.000*	0.942
Visual aids	1.000*	0.838
Live experiment	0.850*	0.823
Audio presentation strategy	1.000	0.889

Table 1: Intercoder reliability.

Note. Human coders completed multiple rounds of training using a sample of 30 videos. We reported intercoder reliability once coders reached a satisfactory level of agreement. The asterisk, *, indicates that we report percent agreement rather than Krippendorff's alpha because the binary (0/1) codes were highly skewed. When one category is rare, Krippendorff's alpha can be unstable and yield misleadingly high or low values.

Dependent and control variables

We operationalized the popularity of interactive science videos using four indicators of audience engagement: the number of shares (M = 227.83, SD = 2,372.62), views (M = 69,187.89, SD = 399,066.80), likes (M = 4,674.65, SD = 31,582.11), and comments (M = 150.40, SD = 697.59). To strengthen the model testing, we controlled for account-level characteristics, including the number of followers, followings, previously posted videos, and verification status, to isolate the effects of communicative features and audio-visual presentation strategies from the account's existing audience base. We also controlled for video length, as prior research suggests that duration shapes engagement patterns on short-video platforms (Lu & Shen, 2023).

Analysis plan

To address **RQ1**, we calculated the proportion of science-related videos published in 2024 using interactive features (stitch, duet, or reply-to). For **RQ2**–**RQ4**, we used Pearson's chi-square tests to compare distributions of communicator identity, communication objective, and audio-visual presentation strategies between interactive and non-interactive science videos.

For **RQ5**, we conducted latent class analysis (LCA) using the poLCA package in R (Linzer & Lewis, 2011). LCA is a model-based clustering approach that identifies latent subgroups from response patterns across multiple categorical variables, making it well-suited for deriving a typology of interactive science videos based on how communicative features and audio-visual presentation strategies co-occur in practice (Rivera et al., 2025). To determine the optimal number of latent classes, we estimated models with varying class solutions (K = 2 to 10) and compared model fit using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with lower values indicating better fit. We selected the four-class solution because it achieved the lowest BIC while maintaining parsimony and being more interpretable than other solutions (see Supplementary Materials E for model-fit statistics and class profiles).

For over-dispersed audience engagement metrics, we estimated negative binomial regression models to examine how video characteristics influenced engagement with interactive science videos (**RQ6**). To reduce the influence of extreme values, we normalized the highly right-skewed distribution of the number of followers, followings, and previously posted videos using the natural log of (x + 1), following prior studies' practices (Bakhshi et al., 2014). Each model included interactive features, communicator identity, communication objective, and audio-visual presentation strategies as predictors, along with all control variables described above. We specified "duet" as the reference category for interactive features, "professionals/experts" for communicator identity, "informing the public about science" for communication objectives, and "speech" for audio presentation strategy. Visual presentation strategies were entered as separate binary predictors since videos could use multiple strategies simultaneously (e.g., a first-person host alongside live experiments and visual aids), and no single reference category was designated for this category.

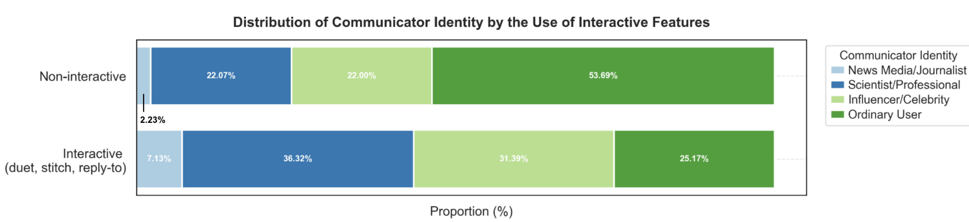
Results

Prevalence and communicative features of interactive science communication

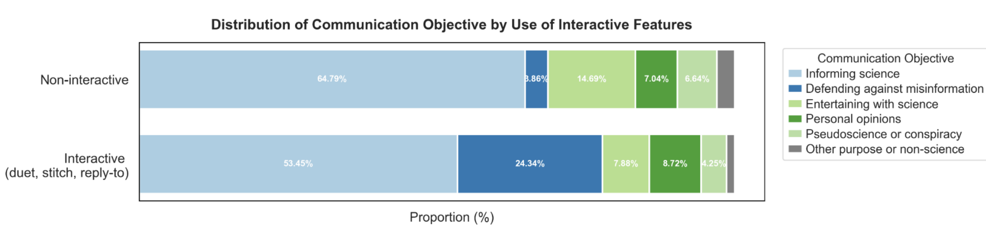
To address **RQ1**, we examined how often TikTok creators used interactive features in science videos published in 2024. In the final sample set for analysis, we identified 1,319 interactive science videos with user-level information and audio tracks, including 49 duet videos, 225 stitch videos, and 1,045 reply-to videos. Overall, interactive formats were used relatively sparingly in science communication: only 3.84% of all science videos in our dataset (N = 34,348) used stitch, duet, or reply-to features.

Regarding **RQ2**, a chi-square test indicated significant differences in communicator identities between interactive ($N = 1{,}319$) and non-interactive videos ($N = 1{,}477$), $\chi^2(3) = 249.87$, $p < .001$ (see Figure 3a). Influencers (31.39%) and scientists (36.32%) were more prevalent in interactive videos than in non-interactive videos (22.00% and 22.07%, respectively). Ordinary users dominated non-interactive videos (53.69%) but comprised a smaller share of interactive videos (25.17%). News media accounted for a small proportion of both interactive (7.13%) and non-interactive videos (2.23%). These results show that professionals and influencers were more likely than ordinary users to adopt interactive features.

For **RQ3**, a chi-square test revealed significant differences across communication objectives between interactive ($N = 1{,}319$) and non-interactive science videos ($N = 1{,}477$) (Figure 3b), $\chi^2(5) = 277.24$, $p < .001$. Defending science from misinformation was more common in interactive videos (24.34%) than in non-interactive videos (3.86%). Non-interactive videos, compared to interactive videos, were more likely to focus on informing the public about science (64.79% vs. 53.45%), entertaining or performing with science (14.69% vs. 7.88%), and promoting pseudoscience (6.64% vs. 4.25%). Personal commentary was expressed at relatively similar rates in both interactive (8.72%) and non-interactive videos (7.04%), and a small share in both groups could not be clearly classified (1.36% and 2.98%).



(a)



(b)

Figure 3: Distribution of communicator identity and communication objective by the use of interactive features.

To answer **RQ4**, chi-square tests revealed significant differences in visual presentation strategies (see Figure 4a): Interactive videos were more likely to feature a first-person host than non-interactive ones (86.13% vs. 42.45%; $\chi^2(1) = 568.65$, $p < .001$). Non-interactive videos were significantly more likely to use visual aids (58.56% vs. 52.16%; $\chi^2(1) = 11.31$, $p < .001$), live experiments (27.83% vs. 15.24%; $\chi^2(1) = 63.84$, $p < .001$), and other formats (9.68% vs. 1.06%; $\chi^2(1) = 96.08$, $p < .001$), compared to interactive videos.

Significant differences were also observed in audio presentation strategies, $\chi^2(3) = 448.48$, $p < .001$ (Figure 4b). Interactive videos relied primarily on speech-dominated audio (80.74%), whereas non-interactive videos used speech-dominated audio less often (41.98%) and more frequently featured music-only (22.48%) or mixed music and speech audio (29.11%) than interactive videos (5.00% and 12.28%, respectively).

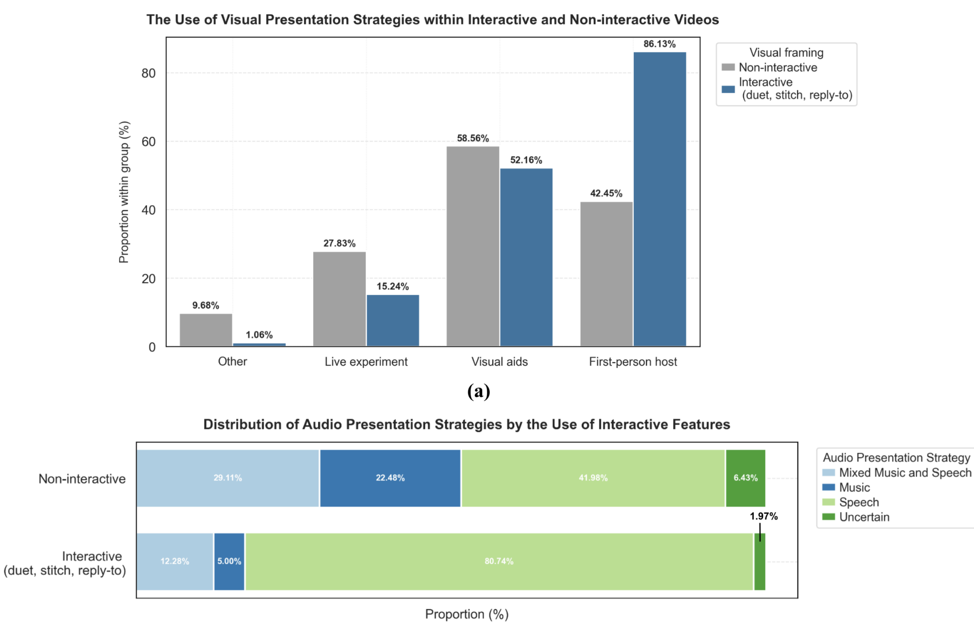


Figure 4: Distribution of audio-visual presentation strategies by the use of interactive features.

Uncovering underlying multimodal patterns of interactive science videos on TikTok

Using LCA, we identified four distinct multimodal patterns that reflect how interactive features, communicator identity, communication objective, and audio-visual presentation strategies co-occurred in interactive science videos (see Figure 5).

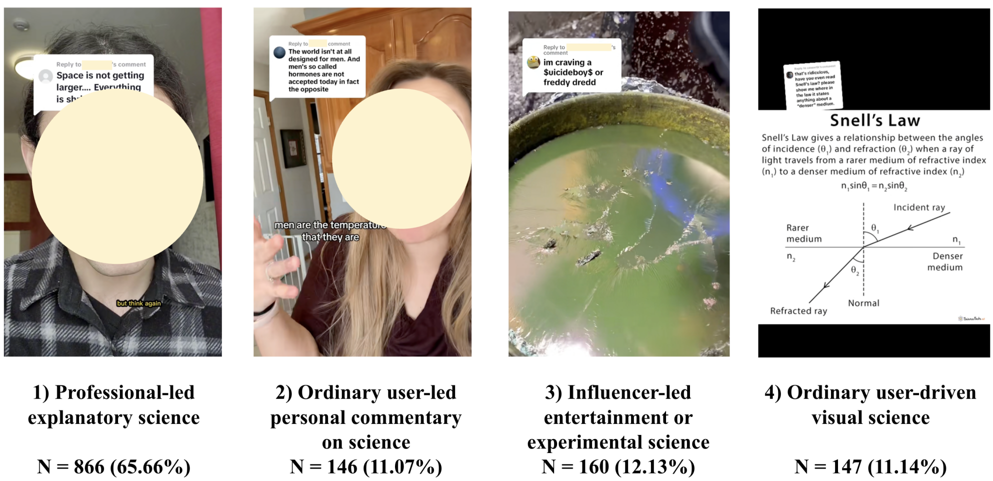


Figure 5: Examples of each co-occurring pattern of communicative and audio-visual presentation strategies.

The dominant pattern is professional-led explanatory science videos (N = 866, 65.66%), which primarily aimed to inform the public about science (66.77%) or defend science against misinformation (28.66%). Over half of the creators had science-related professional backgrounds (51.21%), and most used reply-to (81.67%). These videos featured a first-person host (100.00%), often incorporated visual aids (54.30%), and were largely speech-dominated (92.34%). This pattern reflects an expert-driven, dialogic style in which professionals respond directly to audience questions to inform science. A typical example is a video produced by a creator with a Ph.D. degree, who replied to a user's comment about universe expansion by placing himself at the center of the video and using diagrams to explain complex physics-related principles.

The second pattern is ordinary user-led personal commentary on science videos (N = 146; 11.07%). Creators were mainly ordinary users (48.29%) and influencers (32.87%), often using reply-to (54.52%) or stitch (37.92%) features.

Primary objectives included personal commentary (32.76%), defending science from misinformation (20.33%), and entertainment (17.91%). These videos featured a first-person host (100.00%), occasionally used diagrams or pictures (23.13%), and were mostly speech-dominated (92.30%). This pattern illustrates how interactive features are used to deliver personal perspectives into scientific discourse. For instance, an influencer challenged a user's comment by drawing on their own experiences. In their viewpoint, modern medicine centers male biology, while women's biology is under-researched in clinical trials, eliciting discussions around science and systemic sexism.

The third pattern is influencer-led entertainment or experimental science videos (N = 160; 12.13%). Most videos were produced by influencers (64.31%) and focused on either informing the public (41.16%) or entertaining (38.42%). These creators predominantly used reply-to (89.06%), showcased live experiments (84.40%), and relied on a first-person host to address the audience directly (42.57%). Compared to other classes, this group was more speech-dominated (41.68%) or blended speech with background music (31.53%). An example is a video created by an influencer with over 1 million followers. Responding to a viewer request by growing a bismuth crystal while playing music, the creator transformed a chemical process into a visually pleasing and entertaining performance about science.

The fourth pattern is characterized as ordinary user-driven visual science videos (N = 147; 11.14%). This group was primarily produced by ordinary users (62.93%) and some influencers (24.70%), with reply-to as the dominant feature (87.59%). They typically aimed to inform the public (50.79%) or defend against scientific misinformation (25.50%). This group relied heavily on visual aids (100.00%), and nearly half also included a first-person host (46.34%). Their audio style was more diverse, with 49.76% being speech-dominated and 29.46% combining music and speech. A representative example is a video by an ordinary user explaining how Snell's law mathematically describes how light changes direction when passing between two media with different refractive indices.

Associations between communicative features of science videos and audience engagement

To answer **RQ6**, we fitted negative binomial regression models predicting the number of shares, views, likes, and comments (see Table 2). Regarding interactive features, videos using stitch (β = 0.662, *SE* = 0.263, p = .012) or reply-to (β = 0.651, *SE* = 0.240, p = .007) features received significantly more comments compared to duet videos. However, reply-to videos were shared significantly less than duet videos (β = –0.847, SE = 0.359, p = .018).

Communicator identity plays a crucial role in shaping audience engagement. Influencers (β = 0.460, SE = 0.125, p < .001) and news media (β = 0.473, SE = 0.205, p = .021) received more likes than scientists. Influencers (β = 0.473, SE = 0.115, p < .001) and ordinary users (β = 0.631, SE = 0.169, p < .001) received more comments than scientists. News media received significantly more views than scientists (β = 1.284, SE = 0.226, p < .001).

Communication objectives showed strong associations with engagement. Videos aimed at defending science against misinformation received significantly more views (β = 0.275, SE = 0.122, p = .025), likes (β = 0.363, SE = 0.113, p = .001), and comments (β = 0.954, SE = 0.106, p < .001) compared to videos focused on informing the public. Interactive videos promoting pseudoscience or conspiracy theories received significantly more shares than informational science videos (β = 0.971, SE = 0.351, p = .006), indicating that misleading science content might be particularly likely to diffuse on TikTok. Personal commentary videos received more comments (β = 0.833, SE = 0.161, p < .001) but fewer views (β = –0.481, SE = 0.182, p = .008) than videos informing the public.

Visual presentation strategies were also associated with audience engagement. Videos with a first-person host received more shares (β = 0.585, SE = 0.256, p = .022), views (β = 0.516, SE = 0.183, p = .005), and likes (β = 0.482, SE = 0.180, p = .007). Videos using visual aids (e.g., diagrams, graphics, or pictures) were associated with more shares (β = 0.744, SE = 0.137, p < .001), views (β = 0.275, SE = 0.103, p = .007), and likes (β = 0.286, SE = 0.101, p = .005). Live experiment videos were associated with more views (β = 0.496, SE = 0.168, p = .003). Alternative visual presentation strategies (e.g., animation, scenery, or footage) were not significantly associated with engagement.

Audio presentation strategies exhibited strong associations with engagement. Compared to speech-dominated videos, music-only videos received substantially more shares (β= 1.770, SE = 0.344, p < .001), views (β = 1.576, SE = 0.263, p < .001), and likes (β = 1.445, SE = 0.246, p < .001). Conversely, compared to speech-dominated videos, videos combining music and speech were associated with lower engagement across all metrics: shares (β = –1.068, SE = 0.198, p < .001), views (β = –0.530, SE = 0.152, p < .001), likes (β = –0.662, SE = 0.146, p < .001), and comments (β = –0.693, SE = 0.136, p < .001).

Discussion

This study illuminates the complex realities of interactive science videos on TikTok. First, despite TikTok's affordances for making interactive videos,

	Share Count	View Count	Like Count	Comment Count
(Intercept)	-2.301^{***} (0.689)	6.056^{***} (0.525)	2.509^{***} (0.462)	-1.255^{**} (0.456)
Interactive feature (vs. duet)				
stitch	−0.075 (0.390)	0.410 (0.288)	0.420 (0.274)	0.662^{*} (0.263)
reply-to	-0.847^{*} (0.359)	0.074 (0.265)	−0.469 (0.252)	0.651^{**} (0.240)
Communicator identity (vs. professionals/experts)				
Influencer/Celebrity	−0.133 (0.160)	0.120 (0.134)	0.460^{***} (0.125)	0.473^{***} (0.115)
News media/journalist	−0.473 (0.291)	1.284^{***} (0.226)	0.473^{*} (0.205)	0.141 (0.186)
Ordinary user	0.309 (0.237)	0.029 (0.191)	0.195 (0.181)	0.631^{***} (0.169)
Communication objective (vs. informing science)				
Defending science from misinformation	−0.162 (0.167)	0.275^{*} (0.122)	0.363^{**} (0.113)	0.954^{***} (0.106)
Entertaining with science	0.049 (0.251)	−0.132 (0.204)	−0.196 (0.182)	−0.077 (0.176)
Other purpose or non-science	−0.852 (0.541)	-1.058^{**} (0.393)	-1.136^{**} (0.378)	−0.123 (0.371)
Personal opinions	0.191 (0.233)	-0.481^{*} (0.182)	0.199 (0.176)	0.833^{***} (0.161)
Pseudoscience or conspiracy	0.971^{**} (0.351)	−0.238 (0.289)	−0.496 (0.261)	0.145 (0.239)
Visual presentation strategy				
First-person host	0.585^{*} (0.256)	0.516^{**} (0.183)	0.482^{**} (0.180)	0.296 (0.161)
Visual aids	0.744^{***} (0.137)	0.275^{**} (0.103)	0.286^{**} (0.101)	0.035 (0.090)
Live experiment	−0.112 (0.229)	0.496^{**} (0.168)	0.241 (0.161)	0.091 (0.132)
Other	−0.694 (0.642)	0.029 (0.500)	−0.274 (0.474)	−0.694 (0.436)
Audio presentation strategy (vs. speech)				
Mixed music and speech	-1.068^{***} (0.198)	-0.530^{***} (0.152)	-0.662^{***} (0.146)	-0.693^{***} (0.136)
Music	1.770^{***} (0.344)	1.576^{***} (0.263)	1.445^{***} (0.246)	0.276 (0.220)
Uncertain	0.105 (0.506)	1.270^{**} (0.442)	0.710 (0.382)	0.501 (0.345)
Video length	−0.001 (0.001)	-0.002^{*} (0.001)	-0.001^{*} (0.001)	0.000 (0.001)
Number of followers (log)	1.159^{***} (0.049)	0.882^{***} (0.036)	0.953^{***} (0.036)	0.690^{***} (0.031)
Number of following (log)	-0.090^{*} (0.042)	-0.180^{***} (0.035)	-0.140^{***} (0.032)	−0.005 (0.029)
Verified status (vs. non-verified)	1.830^{***} (0.339)	0.790^{**} (0.259)	0.951^{***} (0.250)	-0.442^{*} (0.219)
Number of historical videos (log)	-0.877^{***} (0.080)	-0.779^{***} (0.062)	-0.784^{***} (0.058)	-0.495^{***} (0.052)
Number of Observations	1,319			
AIC	10263.6	27547.7	20014.3	13018.1
BIC	10388.1	27672.1	20138.7	13142.6

Table 2: Negative binomial regression estimates of the effects of communicative and multimodal features on public engagement. Standard errors in parentheses. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

we found that videos using interactive features were a small share of overall science videos. This may be because interactive videos require creators to invest extra effort in responding to audience comments, navigate platform features, and engage in potentially contentious discussions. This suggests that, as a form of participatory science communication, interactive videos remain limited even on platforms that provide stitch, duet, and reply-to features that encourage the production of interactive videos.

However, among videos using interactive features, we observe patterns suggesting a shift toward participatory science. Diverse communicators were involved in the production of interactive science videos. Scientists and influencers were more prevalent in interactive videos than in non-interactive ones, suggesting that people with professional scientific knowledge or higher familiarity with using social media platforms and managing audiences were more likely to use platform-embedded features. Prior studies have often illuminated the challenges for scientists in changing media environments, such as the slow adoption of contemporary communication practices (Peters, 2013). Here, we found that scientists strategically responded to their audiences' questions using innovative features, pushing back against the traditional image of a "pure scientist" (Walter et al., 2019).

Interactive features indeed enabled varied communication objectives, such as informing the public about science, combating scientific misinformation, entertaining audiences, sharing personal commentary, and even promoting pseudoscience. Notably, interactive videos more frequently defended science against misinformation than non-interactive ones, suggesting these features enable correction rather than mere science education or entertainment. However, the openness also facilitates the circulation of pseudoscience, suggesting the dark side of the participation model (Bucchi, 2008, p. 70). Rather than uniformly democratizing science, TikTok also creates contested spaces where multiple actors use its interactive features for varied purposes, showing opportunities and risks associated with participatory science communication.

Moreover, the audio-visual presentation of interactive science videos reflects a shift toward the personalization of science communication that aligns with the participation model (Zeng et al., 2021). We found that interactive videos primarily featured first-person hosts and used conversational audio. Non-interactive videos relied more on visual aids and live experiments and used background music or hybrid audio formats that prioritized content over the communicator. By foregrounding the communicator with a speech-based delivery, interactive videos shift science communication from an impersonal mode to a more personalized one.

Importantly, scientists constituted the largest share of first-person host videos in our sample (see Supplementary Materials C for details). This suggests that when experts engage with interactive features, they strategically adapt to participatory norms by increasing social presence and approachability (Zhang & Lu, 2023). The communicator-centered presentation fosters parasocial relationships, transforming one-way transmission into dialogues and making scientists appear to be relatable individuals rather than distant figures, thereby facilitating the reciprocity and accessibility highlighted by the participation model.

We identified four distinct styles revealing how different actors appropriate TikTok's interactive features for varied goals: professional-led explanatory science, ordinary user-led personal commentary, influencer-led entertainment or experimental science, and ordinary user-led visual science. These styles illustrate the core characteristic of participatory science communication on such short-form video platforms: diverse actors are involved in science communication and pursue varied objectives beyond one-directional transmission of scientific information to the public. Scientists primarily leveraged the reply-to feature alongside the first-person host presentation and speech-dominated delivery to provide explanations and debunk misinformation, reinforcing their role as educators and epistemic authorities. Ordinary users leveraged reply-to and stitch features with a first-person host and speech-dominated styles to share personal viewpoints, or used the reply-to feature and various visual aids to inform the public about science. Influencers primarily used the reply-to feature and blended live experiments with background music to transform science into visually engaging performances. These stylistic differences illustrate that TikTok's affordances not only enable scientists to maintain expert-driven pedagogy but also simultaneously allow ordinary users and influencers to reimagine science communication through personal expression and visual creativity, producing diverse content within the participatory science paradigm.

Furthermore, engagement patterns reveal tensions in participatory science communication. Influencers and news media received more engagement than scientists, highlighting their prominence on short-video platforms. Videos debunking science-related misinformation generated higher engagement than videos informing the public, indicating that audiences engaged more with contested dialogue than mere knowledge transmission. Most concerning, pseudoscience content received significantly more shares than videos informing the public, suggesting how interactive videos can become a double-edged sword that enables and impedes public science communication.

At the audio-visual level, audiences engaged more with videos featuring first-person hosts, visual aids, or live experiments than other visual cues. Unlike traditional science videos using realistic footage and expository modes (Van Dijck, 2006), TikTok science increasingly shows “science in action” and “scientists behind the lab coats” with various visual informational aids and hosts speaking directly to the camera (Zhang & Lu, 2024). Consistent with prior findings on music in science communication (Crowther et al., 2016), music-only videos received substantially more engagement than speech-dominated ones. However, this effect was not additive: adding music to spoken content reduced engagement, suggesting hybrid audio strategies may impair clarity and processing fluency.

Our study also contributes to computational communication research by extending LLM-assisted content analysis to MLLMs for short-video analysis. While a growing body of work has validated LLMs for text analyses in communication research (Stoll et al., 2025), short-form videos present a distinct challenge: their meaning is constructed through the interplay of visual, auditory, and textual modalities, which traditional machine learning models cannot efficiently capture. We not only demonstrate the feasibility of MLLM-based annotation for a large-scale short-video corpus but also provide a comprehensive framework that incorporates sampling, analysis, and validation steps. However, even when using the same analytic framework, the specific AI model and model version can significantly influence the annotation outputs and may not generate identical annotation results. For this reason, we report the exact models used in this study and encourage future research to revalidate and update the analysis framework when applying it across models and platform contexts.

Overall, our study has both theoretical and practical implications. First, our study contributes to the burgeoning research on participatory science communication by showing that while the platform affords interactive features that could enable the participation model, there is a striking paradox: interactive features remain marginally adopted, yet when used, they afford both participatory possibilities and concerning risks. Second, moving beyond analyzing video-level aesthetic characteristics (Lu & Shen, 2023), we introduce an AI-powered analytic framework for examining communicative and audio-visual presentation strategies in science videos, advancing computational methods for short-video analysis. Third, we constructed a parsimonious typology of interactive science videos and mapped the interplay between interactive features, audio-visual presentation strategies, and engagement dynamics, offering insights into how these elements relate to engagement with science short videos.

Limitations

This study does not come without limitations. First, our multimodal analysis focused on content categories instead of visual-level aesthetic features (e.g., color, brightness) or audio-level acoustic features (e.g., pitch, loudness, and tempo). Future studies could examine how these aesthetic and acoustic features link with audio-visual presentation strategies. Second, communicator identities were analyzed using publicly available information, which may be incomplete or misleading, as creators may not explicitly state or may overstate their expertise. Future research could combine creator surveys or cross-platform verification to enhance coding accuracy. Third, our visual presentation categories, particularly "live experiments," may not fully capture how different disciplines (e.g., the humanities and social sciences) communicate knowledge on TikTok. Future research should develop visual presentation frameworks that account for narrative and interpretive modes of science communication.

Fourth, engagement metrics provide important but limited indicators of audience response. Future research could incorporate audience-level data or experimental design to understand how interactive science videos shape public engagement, knowledge acquisition, and understanding. Fifth, we only measured creators' communication objectives as manifested in their video content rather than relying on self-reported intentions. Future work could complement this approach by interviewing or surveying content creators. Sixth, MLLM-based annotation may introduce biases and is sensitive to prompt design and model choice (Stoll et al., 2025). While we report human validation and provide full prompts, future work should be aware of the potential biases of MLLM-based annotations and validate them across models, languages, and cultures.

Finally, our study focuses on U.S.-published videos with predominantly English content, which limits generalizability in two ways. As GPT-4o is optimized for English, its transcription and classification performance may be less accurate for non-English content. Moreover, findings may not extend to other linguistic, cultural, or platform contexts, as science communication practices vary across countries and platforms. Future research should examine science short videos across diverse geographic regions and platforms.

Data Availability

Replication materials for this study, including analysis code and anonymized data, are publicly available on the Open Science Framework (OSF) at https://osf.io/rwz8q/overview. To protect the privacy of TikTok creators, only anonymized data files are shared; raw video content, visual images, and user profile data are not included in the repository. Video IDs are included in the shared dataset to allow researchers to retrieve the original videos directly from TikTok, subject to their continued availability on the platform.

References

Abidin, C., & Kaye, D. B. V. (2021). Audio memes, earworms, and templatability: The 'aural turn' of memes on TikTok. In C. Arkenbout, J. Wilson, & D. de Zeeuw (Eds.), *Critical meme reader: Global mutations of the viral image* (pp. 58–68). Institute of Network Cultures.

Bainotti, L., & Peeters, S. (2025). Enriching the metapicture: How to analyse audio-visual content on TikTok and other platforms. *New Media & Society*, 14614448251387109. https://doi.org/10.1177/14614448251387109

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 965–974.

Besley, J. C., & Dudo, A. (2022). *Strategic science communication: A guide to setting the right objectives for more effective public engagement*. Johns Hopkins University Press: MD, USA.

Bucchi, M. (2008). Of deficits, deviations and dialogues: Theories of public communication of science. In *Handbook of public communication of science and technology* (pp. 71–90). Routledge.

Chen, K., Kim, S. J., Gao, Q., & Raschka, S. (2022). Visual framing of science conspiracy videos: Integrating machine learning with communication theories to study the use of color and brightness. *Computational Communication Research*, *4*(1). https://doi.org/10.5117/CCR2022.1.003.CHEN

Cheng, Z., & Li, Y. (2024). Like, comment, and share on TikTok: Exploring the effect of sentiment and second-person view on the user engagement with TikTok news videos. *Social Science Computer Review*, *42*(1), 201–223. https://doi.org/10.1177/08944393231178603

Crowther, G. J., McFadden, T., Fleming, J. S., & Davis, K. (2016). Leveraging the power of music to improve science education. *International Journal of Science Education*, *38*(1), 73–95. https://doi.org/10.1080/09500693.2015.1126001

Divon, T., & Eriksson Krutrök, M. (2024). Playful trauma: TikTok creators and the use of the platformed body in times of war. *Social Media + Society*, *10*(3), 20563051241269281. https://doi.org/10.1177/20563051241269281

Dudo, A., & Besley, J. C. (2016). Scientists' prioritization of communication objectives for public engagement. *PLoS ONE*, *11*(2), 1–18.

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Fahy, D., & Nisbet, M. C. (2011). The science journalist online: Shifting roles and emerging practices. *Journalism*, *12*(7), 778–793. https://doi.org/10.1177/1464884911412697

Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, *25*(1), 46–69. https://doi.org/10.1111/comt.12048

Habibi, S. A., & Salim, L. (2021). Static vs. dynamic methods of delivery for science communication: A critical analysis of user engagement with science on social media. *PLoS ONE*, *16*(3), e0248507. https://doi.org/10.1371/journal.pone.0248507

Hautea, S., Parks, P., Takahashi, B., & Zeng, J. (2021). Showing they care (or don't): Affective publics and ambivalent climate activism on TikTok. *Social Media + Society*, *7*(2), 20563051211012344. https://doi.org/10.1177/20563051211012344

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., et al. (2017). CNN architectures for large-scale audio classification. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 131–135. https://doi.org/10.1109/ICASSP.2017.7952132

Huang, C., Fu, T., & Chen, H. (2010). Text-based video content classification for online video-sharing sites. *Journal of the American Society for Information Science and Technology*, *61*(5), 891–906. https://doi.org/10.1002/asi.21291

Jones, A. L., & Stapleton, M. K. (2017). 1.2 million kids and counting—mobile science laboratories drive student interest in STEM. *PLoS Biology*, *15*(5), e2001692. https://doi.org/10.1371/journal.pbio.2001692

Kaye, D. B. V. (2022). Please duet this: Collaborative music making in lockdown on TikTok. *Networking Knowledge: Journal of the MeCCSA Postgraduate Network*, *15*(1).

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge. https://doi.org/10.4324/9780203970034

Lewenstein, B. V. (2003). Models of public communication of science and technology. *Retrieved from https://hdl.handle.net/1813/58743*.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*, 1–29.

Lu, Y., & Shen, C. (2023). Unpacking multimodal fact-checking: Features and engagement of fact-checking videos on Chinese TikTok (Douyin). *Social Media + Society*, *9*(1), 1–16. https://doi.org/10.1177/20563051221150406

Metcalfe, J., Gascoigne, T., Medvecky, F., & Nepote, A. C. (2022). Participatory science communication for transformation. *Journal of Science Communication*, *21*(2), 1–11. https://doi.org/10.22323/2.21020501

Nguyen, H., & Diederich, M. (2023). Facilitating knowledge construction in informal learning: A study of TikTok scientific, educational videos. *Computers & Education*, *205*, 104896. https://doi.org/10.1016/j.compedu.2023.104896

OpenAI. (2025). Introducing GPT-5 [Available at: https://openai.com/index/introducing-gpt-5/ Accessed: 2026-05-09].

Peters, H. P. (2013). Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences*, *110*(Suppl. 3), 14102–14109. https://doi.org/10.1073/pnas.1212745110

Rein, B. (2023). Making science education more accessible: A case study of TikTok's utility as a science communication tool. *Neuroscience*, *530*, 192–200. https://doi.org/10.1016/j.neuroscience.2023.08.021

Rivera, S., Fierro, P., Santos, M., Tagle, F., & Navia, P. (2025). How different patterns of online participation reflect internal, external, and online efficacy: The case of Chile. *Information, Communication & Society*, *28*(8), 1480–1498. https://doi.org/10.1080/1369118X.2025.2452285

Rodríguez Estrada, F. C., & Davis, L. S. (2015). Improving visual communication of science through the incorporation of graphic design theories and practices into science communication. *Science Communication*, *37*(1), 140–148. https://doi.org/10.1177/1075547014562914

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

Stoll, A., Yu, J., Andrich, A., & Domahidi, E. (2025). Classification bias of LLMs in detecting incivility towards female and male politicians in German social media discourse. *Communication Methods and Measures*, *19*(4), 350–368. https://doi.org/10.1080/19312458.2025.2551693

TikTok. (2026). Creator rewards program [Available at: https://www.tiktok.com/support/faq_detail?id=7581821550694013452&category=web_account TikTok Help Center. Accessed: 2026-05-09.].

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, *2006*(146), 93–94.

Trench, B. (2008). Towards an analytical framework of science communication models. In D. Cheng, M. Claessens, T. Gascoigne, J. Metcalfe, B. Schiele, & S. Shi (Eds.), *Communicating science in social contexts: New models, new practices* (pp. 119–135). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8598-7_7

Van Dijck, J. (2006). Picturizing science: The science documentary as multimedia spectacle. *International Journal of Cultural Studies*, *9*(1), 5–24. https://doi.org/10.1177/1367877906061162

Walter, S., Lörcher, I., & Brüggemann, M. (2019). Scientific networks on Twitter: Analyzing scientists' interactions in the climate change debate. *Public Understanding of Science*, *28*(6), 696–712. https://doi.org/10.1177/0963662519844131

Wang, Y. (2020). Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (Douyin). *Computers in Human Behavior*, *110*, 106373. https://doi.org/10.1016/j.chb.2020.106373

Zeng, J., Schäfer, M. S., & Allgaier, J. (2021). Research perspectives on TikTok & its legacy apps| reposting "Till Albert Einstein is TikTok famous": The memetic construction of science on TikTok. *International Journal of Communication*, *15*, 3216–3247. https://doi.org/10.31219/osf.io/8tdvm

Zhang, A. L., & Lu, H. (2023). Scientists as influencers: The role of source identity, self-disclosure, and anti-intellectualism in science communication on social media. *Social Media + Society*, *9*(2), 1–16. https://doi.org/10.1177/20563051231180623

Zhang, A. L., & Lu, H. (2024). Behind the lab coat: How scientists' self-disclosure on Twitter influences source perceptions, tweet engagement, and scientific attitudes through social presence. *New Media & Society*, *26*(10), 5784–5801. https://doi.org/10.1177/14614448221141681

Zimmerman, I., Baram-Tsabari, A., & Tal, T. (2025). Dissemination versus dialogic science communication: How do the deeply involved wish to engage? *Science Communication*, *47*(4), 497–526. https://doi.org/10.1177/10755470241293352