PaperTok: Exploring the Use of Generative AI for Creating Short-form Videos for Research Communication

Meziah Ruby Cristobal*
University of Washington
Seattle, WA, USA
meziah@uw.edu

Hyeonjeong Byeon*
University of Washington
Seattle, WA, USA
hjbyeon@uw.edu

Tze-Yu Chen*
University of Washington
Seattle, WA, USA
alextyc@uw.edu

Ruoxi Shang*
University of Washington
Seattle, WA, USA
rxshang@uw.edu

Donghoon Shin*
University of Washington
Seattle, WA, USA
dhoon@uw.edu

Ruican Zhong*
University of Washington
Seattle, WA, USA
rzhong98@uw.edu

Tony Zhou*
University of Washington
Seattle, WA, USA
tyzhou05@uw.edu

Gary Hsieh
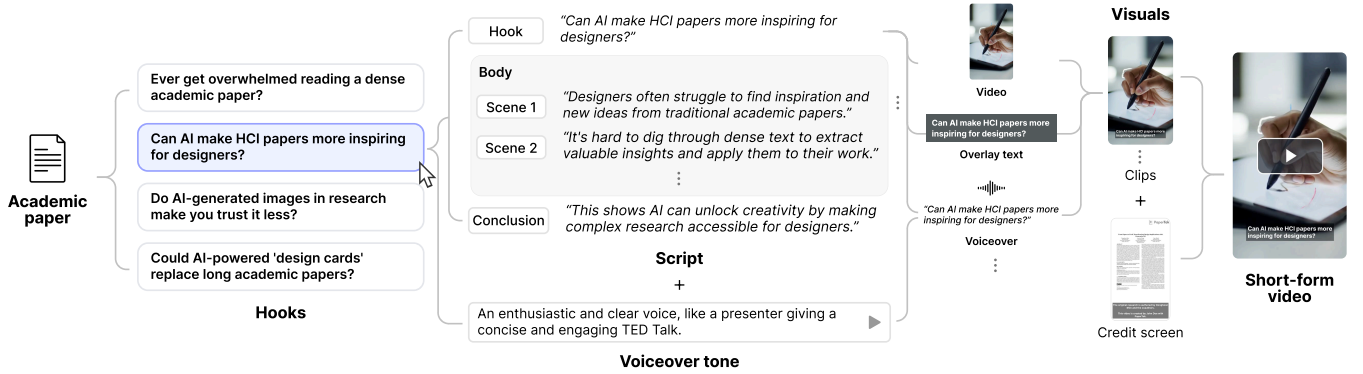University of Washington
Seattle, WA, USA
garyhs@uw.edu

Figure 1: Overview of the PaperTok system, designed to help users transform academic papers into short-form videos through an iterative workflow. Starting with a compelling hook, an accompanying script, and a suggested voiceover style, the process integrates AI-generated video scenes, captions, and a credit screen, streamlining the creation process.

Abstract

The dissemination of scholarly research is critical, yet researchers often lack the time and skills to create engaging content for popular media such as short-form videos. To address this gap, we explore the use of generative AI to help researchers transform their academic papers into accessible video content. Informed by a formative study with science communicators and content creators ($N = 8$), we designed PaperTok, an end-to-end system that automates the initial creative labor by generating script options and corresponding audiovisual content from a source paper. Researchers can then refine based on their preferences with further prompting. A mixed-methods user study ($N = 18$) and crowdsourced evaluation ($N = 100$) demonstrate that PaperTok's workflow can help researchers create engaging and informative short-form videos. We also identified the need for more fine-grained controls in the creation process. To this end, we offer implications for future generative tools that support science outreach.

1 Introduction

The effective communication of scholarly research is a cornerstone of scientific progress and societal well-being. However, the primary medium for scholarly dissemination, the academic paper, is often written in a way that creates a significant *ivory tower* gap between researchers and the broader public. While there are efforts in science communication to bridge this gap, they often require significant time, resources, and a skill set in public engagement that many researchers do not possess [36, 49].

Meanwhile, the digital media landscape has fundamentally reshaped how people consume information. More than 50% of people surveyed in a 2024 Pew report disclosed that they at least sometimes get news from social media [16]. This is further compounded by the meteoric rise of short-form video platforms (*e.g.*, TikTok, Instagram Reels, YouTube Shorts). Characterized by their high engagement, visual storytelling, and algorithmic personalization, these platforms have become dominant channels for information consumption. Pew's 2025 study on TikTok usage found that 17% of adults in the US report they regularly get news from TikTok [45]. Unsurprisingly, researchers and science communicators have begun to use these platforms to share scientific insights [17, 37, 46]. However, creating compelling video content can be a formidable challenge [43]. It demands skills in scriptwriting, video editing, motion graphics, and narration. In addition to the existing obstacles to communicating science [10, 49], these challenges pose a high barrier that prevents researchers from participating in this powerful new medium of public scholarship.

---

*These authors contributed equally to this work.

Recent advances in generative AI have the potential to transform research communication. Large language models (LLMs) can help summarize and synthesize complex information and are already being used for news reporting [40]. Beyond text outputs, generative AI also has the potential to create audio and visual content. This can help facilitate the creation of different types of communication artifacts to support broader access and understandability of scientific content. More recently, with text-to-video models, it is now possible to create video-based content for research communication. This new opportunity for science communication, however, raises important research questions that need to be answered. How can generative AI models be responsibly used to ensure the accuracy of the video and help improve people's understanding and trust in science? How should we design this human-AI collaborative workflow? And, what are people's perceptions of these short-form research communication videos created using generative AI?

While researchers have explored AI-assisted short-form video content creation [47] and identified videos as a promising medium for science communication [14], no prior work has investigated the design and perception of AI-generated short-form videos specifically for research communication. To answer these questions, we began with a formative study with science communicators and content creators. Through a semi-structured interview using AI-generated videos as probes ($N = 8$), we gained insights about how to structure these short-form videos to make them engaging and credible. Using the insights gained, we then developed PaperTok, a novel system that leverages generative AI and employs a human-AI collaborative workflow to empower researchers to transform their academic papers into engaging short-form videos. Researchers upload their paper, and the system uses LLMs to analyze the paper text, identify key findings, and generate a concise, accessible video script. Subsequently, it utilizes a text-to-video model to produce a visual storyboard synchronized with the script. The system presents these AI-generated assets within an interactive editor, allowing the researcher to act as a creative director while ensuring scientific accuracy. This human-in-the-loop approach scaffolds the creative process, combining the efficiency of AI with the essential domain expertise and narrative voice of the human researcher. We developed PaperTok to create a one-to-one translation of academic papers into short-form video to complement existing research communication practices that occur when a paper is published (*e.g.*, social media posts by researchers or university press releases). PaperTok could serve as a new and engaging way to communicate science beyond traditional avenues.

We evaluated PaperTok through a user study with ($N = 18$) researchers who had previously published at least one academic paper. Participants were tasked with creating a short video summary of their work using PaperTok. We also conducted a survey—taken by these researcher participants and ($N = 100$) crowdsourced audience participants—to compare videos generated using PaperTok against videos generated using existing PDF-to-video platforms (*i.e.*, SciSpace[1], PDFtoBrainrot[2]). Our findings indicate that PaperTok is significantly more engaging and entertaining while providing similar levels of informational value. Researcher participants also

described PaperTok as a useful tool for lowering the barrier for generating engaging visuals and voiceovers, while providing for some personalization to suit their communication style. In addition, we uncovered nuanced insights into the human-AI co-creation process, especially about researchers' desired level of control over the workflow and the need to signal the human-in-the-loop in AI-facilitated science communication.

This paper contributes:

- A formative study (§3) that uncovers insights for transforming academic papers into engaging short-form videos.
- The design and implementation of PaperTok (§4), a novel human-AI collaborative system for authoring short-form scientific videos from academic papers.
- Empirical insights from a user study (§5) that demonstrate the effectiveness of our system in reducing barriers to science communication and illuminate the dynamics of human-AI collaboration in a creative, expert-driven context.
- A set of design implications (§7) for future generative AI tools that aim to scaffold, rather than automate, complex creative tasks—emphasizing the need for user control, iterative refinement, and the seamless integration of AI suggestions with human expertise.

2 Related Work

We outline two areas of prior work that we build on: (1) participatory web and the use of short-form videos for research communication, and (2) the potential use of generative AI for science communication.

2.1 Participatory Web for Research Communication

Much research has examined the use of participatory web to communicate science [35, 39]. This includes the use of blogs [18, 29], social media [9], video-sharing sites [48], podcasts [52], etc. These platforms have disrupted the traditional publication pipeline [42, 43, 49] and empowered researchers to be content creators and communicate their scientific findings with the lay public directly [32]. For example, researchers have used science blogs to re-contextualize research findings to be more personally relevant to readers [29], forums like r/science to facilitate scientific discussion forums like r/science [23], and social media like Twitter to foster public engagement with science [8, 22]. Not only are these efforts valuable for educating the lay public and improving society's overall science understanding, but they are also valuable to researchers themselves. Researchers communicating research on the participatory web can improve the visibility of their work and increase citation counts [26], as well as receive feedback from peers and collaborators [9].

Video has been an important format for science communication, as combining narrative content with compelling imagery increases audience immersion and message effectiveness [14]. As people increasingly consume short-form videos for entertainment and information, these platforms also become increasingly important channels for science communication [45]. Despite this, studies about research communication on TikTok remains limited and typically focuses on individual accounts [37, 53]. For example, one study of

---

150 neuroscience TikTok videos from @dr.brein found that shorter videos minimized viewer attrition, and that videos summarizing research articles outperformed those offering advice [37]. A different study analyzed just one account again—@TerraExplore, focused on geoscience and geophysics—showed that its TikTok videos generally received more views compared to their cross-posted videos on YouTube (Shorts) and Instagram (Reels), highlighting TikTok's algorithmic advantage for reaching audiences [53].

However, while social media can help researchers communicate broadly, the effort required can still pose barriers for researchers. Most researchers simply do not have the time to do science outreach [12, 49]. Even if they have the time, they may lack training in translating technical jargon into plain language or in presenting their work in ways that engage diverse audiences [10, 21, 49]. These challenges are now compounded when attempting to communicate through short-form videos, as researchers report lacking the time and skills to develop public-facing audios and visuals [43]. Combined, these challenges limit scientists' participation in short-form video science communication.

2.2 Rise of Generative AI and its Role in Science Communication

Many scholars have argued that the rise of generative AI can have a transformative impact on science communication [1, 3, 38]. Interviews with HCI scholars and other scientists have also revealed that generative AI tools are being used across research practices [24, 31]. Text-to-image models are used to create visuals for more engaging and understandable artifacts to communicate design implications from research findings [41]. Text-to-audio models are used to create podcasts of scholarly papers (*e.g.*, NotebookLM's Audio Overview[3]). Other studies have explored the use of LLMs to generate content ideas [13], translate scientific jargon to laypeople [2], and transform papers into lyrics for performable karaoke tracks [7]. Similarly, prior research on text-to-video models has examined their potential for education [30] and for patient outreach and medical training [44], but not for science communication. Moreover, while short-form video is becoming increasingly influential as a medium for public engagement, its value as a science communication format—and the role of researchers in this process—has not been systematically studied.

Commercially, there are platforms exploring the conversion of research papers in PDF format into videos. For example, SciSpace transforms them into video summaries, while PDFtoBrainrot converts them into "entertaining, TikTok-inspired *brainrot* content" to maintain attention and therefore facilitate learning. However, neither of these tools allows for collaborative input from users during video creation. It is also unclear how people perceive these generated videos. This raises questions about the ability of these existing PDF-to-video services to support science communication, as well as how to design for researcher-AI collaboration in this context.

Furthermore, despite the potential benefit generative AI holds for research communication, several concerns have also been noted [38]. First, there are concerns about the accuracy of using generative AI to support science communication. While continuous advances in

LLMs have improved on the limited quality noted in earlier models [34], more current models are still susceptible to hallucination and logical fallacies [27]. LLMs are also known to exhibit biases [28]. These issues can be especially problematic in the context of science communication in high-stakes domains where precision matters. Concurrently, researchers have also expressed concerns about how generative AI may contribute to the growing 'infodemic' problem—where people are overloaded with excessive information [11]. The lay public may experience increasing difficulty discerning between high-integrity information when overloaded with AI-generated content. These concerns highlight the need for critical inquiry into AI's potential role in short-form science communication and for guidelines that support accurate, trustworthy content.

3 Formative Study

To inform the design of PaperTok, we conducted a formative study with science communication experts and content creators. Our goal was twofold: (1) document experts' current practices (*i.e.*, how they plan, script, and produce short-form science communication videos), and (2) elicit their perceptions of what makes these videos engaging, understandable, and credible for target audiences. Together, these perspectives surface design requirements for tools that transform academic papers into short-form videos.

3.1 Participants

We recruited 8 participants (Table 1) through targeted outreach to science communication content creators from Instagram, RedNote, TikTok, and YouTube, as well as university communication experts. Eligible participants were adults who create, edit, or manage original science communication for public audiences as independent creators, institutional communications staff, or researchers. We required recent original science-focused work and the ability to explain how they source and translate academic work; we excluded accounts that only reposted others' content or were inactive channels.

3.2 Procedure

We conducted 60-minute semi-structured interviews via video call, and each participant received $30 USD in their choice of a gift card. Each session began with questions about participants' content creation workflows, challenges in translating research, and experiences with AI tools. In the second half, participants reviewed science communication video probes, evaluated their hooks, scripts, and visual concepts, and reflected on what they perceived as effective or problematic from an audience standpoint. We concluded by co-designing improvements with participants. They rewrote hooks, suggested visual alternatives, and identified missing narrative elements. Sessions were recorded, transcribed, and analyzed using thematic analysis to identify recurring patterns in quality criteria and design requirements.

3.2.1 Video probes. For our video probes' source papers, we selected three CHI 2025 Best Paper awardees that each represented a different HCI research contribution [50]—artifact [25], empirical [20], and methodological [33]. We converted each paper into a video using three methods: (1) generating a script with Google's Gemini 2.5 Flash and converting each script segment to a video

---

Table 1: Formative study participant demographics. P1, P4, and P7 held institutional roles managing content across multiple outlets rather than maintaining singular representative accounts, so we indicated videos posted as not applicable (N/A).

PID	Job title	Main platform(s)	Scicomm experience	Videos posted (past year)
P1	Communications Manager	University and departmental news channels	12 years	N/A
P2	Full-time Content Creator	YouTube, TikTok	5 years	25-49
P3	Part-time Content Creator	RedNote	1 year	100+
P4	Digital Content Creator	University news outlets	2.5 years	N/A
P5	HCI Researcher & Faculty	TikTok	5 years	100+
P6	Full-time Content Creator	TikTok, YouTube, Facebook, Instagram	6 years	100+
P7	Director of Content & Communications	University news outlets	15 years	N/A
P8	Part-time Content Creator	RedNote	1 year	50-99

clip with Veo 2 (`veo-2.0-generate-001`), then compiling them into a single short-form video; (2) using PDFtoBrainrot, and (3) using SciSpace's PDF to video service. Each video was generated by the same research team member to maintain methodological consistency and enable a controlled comparison between the three generation approaches.

PDFtoBrainrot videos consist of colorful visuals, showcasing playthroughs of popular games such as Subway Surfers and Minecraft. The scripts use slang words such as *"Skibidi bop"* for humor, and captions are shown one word at a time as a bid for prolonged attention. By contrast, SciSpace primarily uses the reference paper's figures on a white background for its visuals. Its scripts are scholarly summaries, and captions are shown a few words at a time, with yellow highlights timed with the narration. Because of these stylistic differences, we chose PDFtoBrainrot as a comparator for engagement, and SciSpace as a comparator for informativeness.

3.3 Formative Study Findings

In general, participants reported a desire for tools to increase their capacity. As P2 said, *"the largest challenge is just time."* Interestingly, several participants reported already using LLMs to assist with background research, and some had experimented with the use of AI-generated content in their videos as supplemental material.

When it came to effective short-form science communication videos, participants reported several design decisions where narrative, visual, and audio elements worked together to maintain engagement and credibility. To reflect this interdependency, we present our findings as key design challenges that cut across various video elements:

- Script: The narrative structure that translates academic findings into accessible language
  - Hook: The opening 2–5 seconds that captures immediate attention through questions, surprising facts, or relatable pain points
  - Body: The main narrative that presents research context, key findings, etc.
  - Conclusion: The final part of the video that provides actionable takeaways and explicitly resolves the initial hook's opening question and creates narrative closure
- Audio: The vocal narration that delivers the script with voiceover
- Visual: The imagery that illustrates the script through video clips, embedded figures, etc.

Participants' workflow reasoning and assessments of video quality were tightly linked. For this reason, we report our findings as four cross-cutting design challenges that integrate both practices and perceptions, a structure that more directly surfaces design requirements for PaperTok.

3.3.1 Challenge 1: Which research content should be communicated? Effective science communication videos must strategically select and frame research content to resonate with viewers' lived experiences while maintaining utility or educational value. This is foundational to the video generation as the initial framing stage that determines what content should be presented.

Personally relevant. The content needs to have an impact on viewers' daily lives. As P8 puts it, *"How does this impact me as a person, and where am I? And like, what does this mean to me in my daily life?"* This focus on personal stakes addressed a communication challenge, as P1 observed: *"People are saturated with information all the time. And so how do we get them to care about this?"* According to participants, rather than presenting abstract findings, effective videos needed to demonstrate concrete problems and solutions. P1 emphasized, *"I want to show how this is a problem for people. And I want to show what our research is doing to help."* P4 similarly prioritized identifying, *"what is the most important aspect of this in relation to like the user or just the general person."* In addition, P1 expressed that communicating the 'why'—'Why did the researcher find this to be something that they wanted to study?'—would also be personally relevant to the audience, *"because that's really going to tell people why this is something they should care about."*

Timely. Participants sought material that addressed *"what's the latest"* (P1) and aligned with trending topics. P4 considered whether it is a topic *"that people are currently talking about."* P9 selected content *"close to a current hot topic, like AI's inconsistent goals in local versus national government."* P1 recommended coordinating with news cycles: *"If there's something that is kind of in the news... you should check out what our researchers are doing."* Even specialized research needs everyday relevance. P6 was *"always trying to look for hooks or things that make people understand why the things we do in space matter back here."*

3.3.2 Challenge 2: How can immediate attention be captured? The opening seconds strongly influence whether viewers continue watching or scroll past: *"The issue with short-form social media is that someone will decide in 2 seconds whether they're gonna keep watching it or scroll past it."* (P5)

Hook design. Effective hooks are crucial, as ineffective hooks *"took too long to get into what the topic was about"* (P1), burying the main point. We identified the following principles for effective hook design:

Effective hooks acknowledge common pain points. P2 praised one of the hooks from the videos we showed—*"Did you ever feel like your settings don't matter"*—noting it *"speaks to me as a consumer of social media."* Universal experiences—such as getting interrupted while working—also provided instant recognition, helping viewers engage with the science content that followed.

Hooks should generate 'wait, what?' moments that require resolution (P6). P3 uses hooks that address misconceptions: *"Depression and constant travel aren't contradictory - you can be depressed in Bali."* P2 described subverting expectations by *"taking something that is used in a very traditional manner, and then putting it in a very different situation."* Unexpected statistics also grabbed attention—after they were shown the example hook of *"98% of Nigerians distrust police,"* P6 reacted *"98, that's a really high number, and that's why I gravitate towards that."*

Hooks must be *"short and punchy, but you also don't want to lack context"* (P6). Effective creators started general before going deeper, *"trying to keep it pretty general at the beginning...knowing [they're] going to have time to go more in depth later"* (P6).

Visual and audio quality for immediate engagement. The participants universally dismissed SciSpace video's primarily static screenshots of figures from the paper source. P5 stated bluntly: *"Just [having] the image of the paper is, you know, anyway, no one's gonna watch that."* P2 compared it to the other video probes, stating: *"I think [dynamic visuals were] more engaging than the previous one, which, just like, showed a picture of the research paper, which I don't think is exciting for anybody."*

In terms of audio, artificial-sounding narration triggered instant rejection. P3 reacted to the video generated by SciSpace: *"This is terrible. The AI voice immediately lowers my interest."* P8 emphasized the need to *"humanize the content"* through narration. The opening tone needs to establish energy and authenticity immediately, with P1 noting that just hearing an 'AI voice' made them think that: *"God, no one even worked on this... like this is just a mess."* The tone should also match the energy of short-form content. To P3, this meant avoiding academic presentation styles that was akin to *"listening to a boring academic presentation."*

Participants also suggested using text overlays that reinforced hook questions or statements, enabling viewers to process information through multiple channels simultaneously. P9 noted that *"subtitles highlighting spoken words make it engaging and less tedious,"* while P4 observed that *"even just having video with text over it can be effective."* This was particularly important given platform viewing behaviors where videos often autoplay without sound. However, P3 provided perspective on their relative impact: *"Content is king. [Caption] highlighting accounts for maybe 1% of impact. Good content makes people stay even without subtitles."*

3.3.3 Challenge 3: How can engagement be maintained throughout? While hooks capture initial attention, sustaining viewer engagement throughout the video presents distinct challenges.

Forming a fast-paced, complete narrative. Supporting prior findings [37], participants consistently identified shorter video duration as crucial for retention. P4 explained: *"15 to 30 seconds is really what people are going to pay attention to. So anything beyond that, you potentially lose interest."* While some participants extended this to *"under 60 seconds"* (P4) or *"around 2 minutes long"* as an absolute maximum (P3), the consensus favored the shortest possible format that could still convey meaningful information. Beyond duration, effective videos required a clear narrative that fulfilled their opening promises. Participants emphasized that conclusions must circle back to resolve the initial framing. P2 noted: *"I would almost expect, at the ending, for it to have some conclusion for me as the same person who was talked to at the beginning."* This narrative closure needs to be actionable rather than abstract. P8 suggested adding *"some sort of conclusion, (...) like, so what does this mean for them? What are the next steps? What can they do?"* Thus, the challenge is in creating a complete story—from problem to findings and practical implications—within a short duration.

Making complex research understandable. Participants identified challenges in making abstract academic concepts accessible within short-form constraints. P2 articulated a fundamental difficulty in communicating research that involved subjects that are not easy to visualize, such as *"something subatomic, or if you're talking about some conceptual thing."* To address this, visual translation emerged as a potentially useful strategy for comprehension. P9 suggested using animations, stating they *"would be very helpful, especially for visualizing abstract concepts or data from the paper. It would make the content more engaging and easier to understand. For example, if there's a flow chart or a process, animating it would be much better than just showing a static image."* Beyond visual aids, participants emphasized connecting unfamiliar concepts to familiar experiences. P3 demonstrated this approach when discussing privacy research: *"Connect it to relatable examples. Hidden cameras, hotel spying, seeing yourself on adult websites. Japanese iPhones can't turn off camera sounds because of anti-spying measures. People hate being secretly filmed."* This strategy transformed abstract privacy concerns into immediate, emotionally resonant scenarios that viewers could instantly understand.

Maintaining interesting visuals. Participants wanted visuals that showed *"actual results of the study"* rather than generic stock footage (P2), though achieving this specificity proved difficult. P2 observed that while some visuals were *"content related,"* they were not *"really visually that interesting."* The challenge was particularly acute for abstract research, with P9 noting: *"Many papers lack a clear theme for visuals, so I have to search for images that are abstract yet relevant. This takes a lot of time."* When screenshots of academic papers appeared on screen, participants emphasized they needed a dynamic presentation rather than a static display. P4 suggested visuals should be *"zoomed in on certain parts that it was referring to, or highlighted certain, like, interesting things"* to maintain visual momentum. The overarching need was for *"something visual that... [can] be told through a video format."* (P6)

3.3.4 Challenge 4: How can credibility be communicated while remaining engaging? Participants identified credibility as crucial for

science communication videos, particularly given growing skepticism about AI-generated content.

Adding human authenticity signals. Human presence emerged as the primary credibility signal. Participants highlighted successful creators like Cleo Abram who *"does the whole voiceover, but the first half second is always her face in every single video"* (P2). This brief personal appearance was enough to establish trust for P2: *"Even if I don't know what this topic is going to be about, I know that Cleo talks about interesting science stuff, and so I'm more willing to stick around."* Even having a human read AI-generated content *"would make a world of difference. It would create more credibility. It would make it more human"* (P5). In addition, participants valued natural imperfections. P2 noted: *"I would fumble with my words (...) [but] that just makes it feel like it's not just a robot-generated spew."*

Providing academic authority markers. Participants suggested adding explicit citation, such as authorship or institutional branding (P1). P5 noted their absence in the PDFtoBrainrot video probes with dismay: *"There's still no credibility markers in this, and it doesn't even cite the paper."* Moreover, visual elements needed to reinforce academic credibility. P9 stated: *"Simple, abstract, and professional designs are preferred."* P2 distinguished between *"visuals and visuals that add to the story,"* noting that random imagery without connection to the content diminished credibility.

Avoiding credibility pitfalls. Artificial visuals particularly undermined credibility by signaling low effort and a lack of care. P1 observed: *"It just looks overwhelmingly AI, which maybe that's okay. But, to me, when I see that I have a sense of somebody didn't spend time on it... that [it] feels kinda, like, careless."* In addition, the robotic quality of AI voices, characterized by *"lack of tone variation, absence of pauses, and words that almost mashed together"* (P4), immediately signaled low production effort and reduced viewer trust. P5 predicted this sensitivity would intensify: *"People are going to become even more sensitive to AI slop because there's gonna be more and more of it to the point where, like, Oh, my God, it's a human! Thank God! Let me watch this."*

4 The PaperTok System

Building on our formative work, we designed PaperTok, an interactive system that streamlines the content creation process while supporting user creativity and information credibility and accessibility. In this section, we describe the user flow interacting with PaperTok, as well as each feature's design considerations and implementation.

4.1 User Flow

PaperTok's workflow guides users from content ideation to video production. LLM is used to quickly generate content from the research paper, while human-in-the-loop checkpoints ensure human users maintain creative control.

The process begins once the user uploads a PDF of their academic paper, and they are taken to *Step 1: Hook + Script* (Figure 3). On the left panel are four hook options with matching scripts. On the right panel is an editor where the currently selected script is displayed, so users can refine their narrative. Below these panels are AI-driven recommendations for voiceover tones for each script, which users can preview and personalize by editing the prompt.

In *Step 2: Storyboarding* (Figure 4), users are presented with a potential script as segmented scenes, each paired with a corresponding visual to be generated. This design affords flexibility: users can generate all scenes at once then review and refine, or they can work scene by scene, refining a portion of the script and generating its visual before moving on. Users retain control throughout and engage in an iterative, storyboard-workflow, where they can continually revise scripts and regenerate visuals until all scenes are satisfactory.

In the final step, *Step 3: Credit Screen* (Figure 5), users are prompted to add their name in a created by-line alongside author attribution for the uploaded paper. This Credit Scene is appended to the end of the video, signaling credibility of content and the human-in-the-loop workflow.

4.2 Script Generation

Instead of generating a monolithic script from the start, PaperTok prioritizes the selection of a compelling hook, which guides the generation of a script that is thematically coherent and logically flows from this hook. In this section, we detail the design and implementation of this process.

4.2.1 Hook generation. We designed a multi-step pipeline that systematically transforms academic findings into compelling hooks. This process is implemented through a system prompt that guides the LLM through a structured sequence of analysis, ideation, and refinement. The pipeline begins by instructing the LLM to identify and extract four core findings that are concrete, relatable, and have clear real-world impact. This initial step directly addresses the design requirement to be *"personally relevant"* by forcing the model to prioritize the paper's most tangible contributions over more abstract or theoretical ones. This ensures the generated content can be engaging to audiences.

Next, the LLM is prompted to brainstorm four distinct story ideas by applying a set of predefined creative tactics derived from our study's findings on creating cognitive tension. These tactics include leveraging contradiction (*e.g.*, pitting the research against a common belief), highlighting surprise or irony, establishing personal stakes (*e.g.*, *'Will this replace YOUR job?'*), and creating curiosity. From a technical standpoint, this structured brainstorming prevents the LLM from defaulting to generic summaries and instead encourages it to explore multiple engaging angles for the same piece of information. For each story idea, the LLM then generates a short, compelling narrative description, which forms the basis for the final hook.

The final stage focuses on crafting punchy, clear, and engaging hooks. Here, the prompt imposes a series of strict stylistic and safety constraints. Each hook must be conversational, without academic jargon, and limited to a maximum of 15 words to ensure brevity. A key technical consideration is the instruction to convert definitive statements into questions (*e.g.*, *'AI causes X'* → *'Could AI cause X?'*). Through this, we aimed to mitigate the risk of oversimplifying or misrepresenting research findings, while enhancing viewer engagement by posing a question that the rest of the video promises to answer. Finally, the system includes a quality control step where the LLM is prompted to review and rate all generated hooks for

Table 2: Formative study findings and its implications to the design and implementation of PaperTok

Challenge	Design implication	PaperTok implementation
Which research content should be communicated?	Surface claims that could be personally relevant and timely	Extracts concrete findings and generates four hook & script options with engaging angles
How can immediate attention be captured?	Include short, punchy hooks; ensure strong visuals and voice	Produces ≤15-word hooks framed as engaging questions; suggests relevant voiceover tones
How can engagement be maintained throughout?	Create fast-paced narratives with relatable examples and visuals	Generates 8-scene scripts that follow a complete problem-solution arc, shown to the user in a storyboard-based UI
How can credibility be communicated while remaining engaging?	Surface human authenticity and credibility signals	Ends video with screenshot of paper and editable, auto-filled attribution that creators can add their names to

engagement, relevance, and emotional appeal. Only the top four curated options are presented to the user. This ensures the output is not only creatively diverse but also pre-vetted for effectiveness.

4.2.2 Script. Once the hooks are generated, PaperTok proceeds to generate the full video scripts for each hook candidate. A central finding from our formative study was that a hook's effectiveness is tied to the narrative it promises; a great opening line that leads to an unrelated or confusing story results in viewer disappointment and disengagement. To address this, PaperTok generates a complete script for each of the four candidate hooks. This process is implemented through a detailed, multi-step LLM system prompt that guides the model to produce a structured, conversational, and visually-oriented narrative.

First, the narrative of each script is structured into eight scenes—what we found to be a good segmentation for a short-form video so that transitions are not too quick, but still fits within the typical short-form video duration of around 45 seconds. Hook (Scene 1) introduces the premise or question. The Body (Scenes 2–7) develops the story, which consists of context (Scenes 2–3; presenting the problem or setting in a simple, relatable way), findings (Scenes 4–5; revealing the core insight, often through a visually intuitive metaphor or example), and relevance (Scenes 6–7; explaining the real-world significance, answering the audience's "so what?"). The Closing (Scene 8) completes the narrative loop, answering the initial question or explaining the surprising claim. This approach ensures a clear takeaway aligned with participant feedback. Each scene is limited to 18–22 words (6–7 seconds), supporting attention and distinct visual storytelling.

Second, we prompted the model to shape language to be both credible and accessible through a set of rewriting rules embedded in the prompt: academic jargon is converted into everyday language, complex sentences are restructured into conversational phrasing (*e.g.*, *"The findings suggest..."* → *"This means..."*), and abstract terms are replaced with concrete, relatable examples (*e.g.*, *"cellular devices"* → *"your iPhone"*). To address the challenge of visualizing abstract concepts, the prompt encourages the LLM to either use tangible real-world analogies or describe concepts in ways suitable for animation. These instructions directly respond to participant feedback requesting content that feels human, expressive, and free of robotic, unemotional delivery.

Finally, the prompt embeds a self-correction step, where the LLM is instructed to review its own output against a checklist. This internal review assesses the script for coherence, conversational quality,

abrupt endings, and any lingering jargon or inaccessible language. This quality assurance step ensures the final script presented to the user is polished and aligns with the platform's stylistic norms. The final output for each of the four hooks is a structured JSON array, where each object contains the scene index, script text, and a calculated duration. This machine-readable format seamlessly integrates with the subsequent voice synthesis and video generation stages of the PaperTok pipeline.

4.2.3 Content-aware voice tone. PaperTok employs a customizable, two-stage voice generation process designed to balance platform-native engagement with content-specific appropriateness. Based on our formative study's findings that short-form video viewers expect a high-energy pace, we defaulted the voiceover tone as: 'an influencer vibe with fast speech.' This baseline ensures that all generated audio possesses a fundamental level of energy suitable for the medium, preventing the output from sounding overly dry or academic, which participants found disengaging.

Recognizing that a single tone cannot adequately represent the diverse emotional landscape of academic research, PaperTok proactively analyzes the script's narrative content, mood, and key message to provide a recommendation for a stylistic modifier. This is implemented in the LLM routine, which leverages an LLM to analyze the script and suggest a single, descriptive line that best complements the story (*e.g.*, a paper on the detrimental effects UI dark patterns → *'with an urgent, cautionary tone'*).

The user is then presented with this AI-generated recommendation within an editable text field, placing them in control of the final output. They have the agency to accept the suggestion as is, refine it to better match their vision, or discard it entirely and write a new style prompt from scratch. Once the user finalizes their stylistic preference, the LLM routine is triggered to combine the voiceover style. This client-side function takes the user's input and uses an LLM to merge it with the baseline voiceover style. The LLM is specifically instructed to preserve the core 'fast speech' directive, while integrating the user's desired emotional and tonal characteristics, creating a final, consolidated prompt. For instance, if a user requests a *'more serious and authoritative'* tone for a script about policy implications, PaperTok converts it into *"Speak fast with a serious and authoritative vibe."* This consolidated prompt is then passed to the backend, which prepends it to the script text before sending the complete payload to the text-to-speech model.
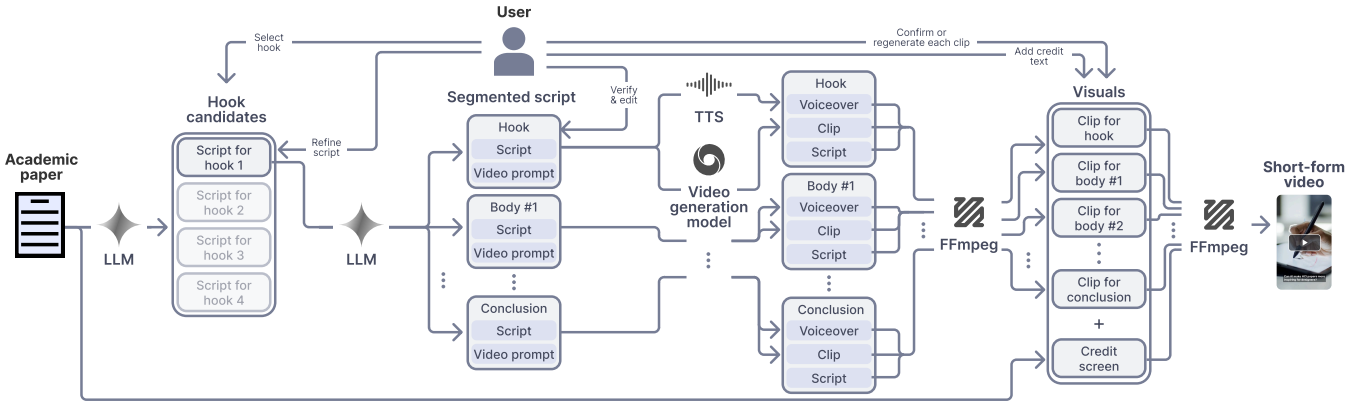
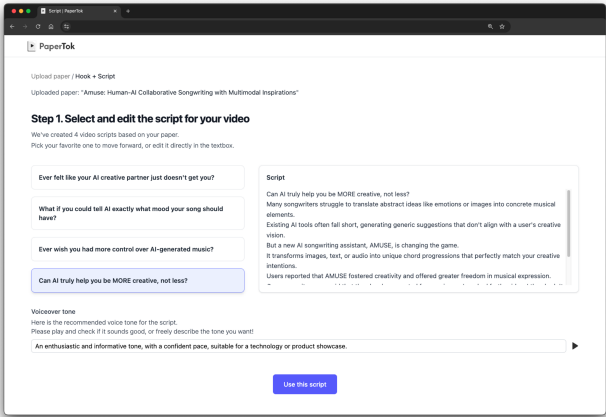**Figure 2: Overview of the PaperTok system's technical pipeline and the user interaction**

Figure 3: PaperTok's *Hook + Script* step, where the user selects and edits their hook, script, and voiceover tone

Figure 4: PaperTok's *Storyboarding* step, where the user enters into an iterative visual generation workflow

4.3 Video Creation

Once the user selects a hook and finalizes the script, the workflow transitions from narrative design to visual production. This stage is meticulously designed to balance automated efficiency with fine-grained creative control, allowing users to iteratively shape the final video on a scene-by-scene basis. The process is architected as a multi-step pipeline, beginning with automated preparation, moving into an interactive user-driven generation phase, and culminating in the final assembly of the video.

4.3.1 Script segmentation & video prompt generation. Before the user begins generating visuals, PaperTok performs two preparatory steps in the background. First, to account for any edits, additions, or deletions made by the user in the script editor, an LLM routine re-segments the final text into discrete scene units. This ensures that the video structure remains coherent and that each segment corresponds to a manageable narrative beat, even if the user has deviated from the initial eight-scene template.
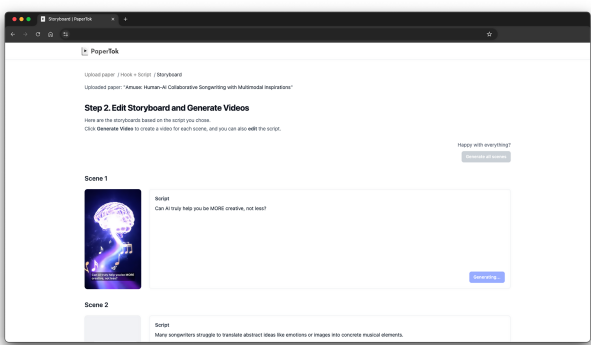
In parallel, another LLM routine generates a detailed visual prompt for each newly defined segment, converting the semantic content of the script into a set of instructions tailored for the video generation model. The prompt is generated to be highly descriptive, specifying not just the subject matter but also the desired mood, style, and camera action. For instance, PaperTok translates a script line of *"This means your phone is always listening"* into *"A stylized, cinematic animation of abstract sound waves flowing from a person's mouth towards a glowing smartphone on a nightstand, dark and moody atmosphere, subtle dolly zoom."*

4.3.2 Generating each clip. With the segments of the script and their corresponding visual prompts prepared, the user enters an interactive generation panel, which presents a storyboard view (Figure 4). Each segment of the script is displayed with its corresponding script text and an option to generate the visual clip. When a user initiates generation for a single scene, the frontend first sends the visual prompt, script text, and customized voice style to the backend, which immediately assigns the request a unique job identifier.

Then the backend process conducts the generation, where it first synthesizes the voiceover for the scene using the provided script and voice style. The server then calculates the duration of this
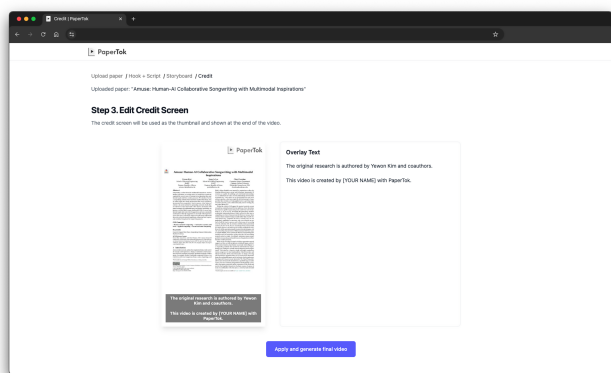
Figure 5: PaperTok's *Credit Screen* step, where the user is asked to sign-off on their creation and can add additional attribution details

resulting audio clip, which is used to dynamically set the required duration for the video clip generation. The video generation model is then invoked with the visual prompt and this exact duration, with a half-second of padding to avoid an abrupt transition to the next clip.

After the raw video is successfully generated and downloaded to the server, the system uses FFmpeg[4] to merge the silent video with its corresponding voiceover track. It also overlays the scene's script as burned-in text captions, a feature that improves accessibility and aligns with the established conventions of short-form video platforms.

4.4 Credit Scene as Credibility and Human-in-the-Loop Signals

Drawing on formative study findings that visible human attribution enhances credibility, PaperTok concludes with a credibility-building module (Figure 5).

To construct this "credit scene," PaperTok programmatically extracts a screenshot of the paper's first page and the names of the authors. The screenshot is used as the visual link to the original academic work, and the author names are pre-populated in the attribution statement—*"The original research is authored by <author names>. This video is created by [Creator Name] with PaperTok."*—that is overlaid on top of the screenshot.

This attribution statement serves the purpose of giving clear, formal credit to the original researchers, while transparently identifying the video's creator, the tool used, and the human-AI co-creation workflow. Because users retain full flexibility to modify this text, they can also add their own signature, social media handle, or institutional branding, allowing the final video to foreground the human creator within an otherwise AI-forward workflow.

4.5 Merging

Once all individual scene clips have been generated and the credibility screen is finalized, the user triggers the final merge by clicking

a button in the interface, signaling the completion of their creative process (Figure 5). This action sends a list of approved video snippets and the credibility screen data to a backend endpoint, which initiates an asynchronous merging task. The static credibility image is first converted into a two-second video clip with silent audio, and then FFmpeg's concatenation filter stitches together all scene clips and the credibility-screen video in the correct order, combining both video and audio streams into a single MP4 file. Once processing is complete, PaperTok provides the user with a preview and download link to their completed short-form video.

4.6 Implementation

PaperTok is implemented as a web interface, where users can interact with generative models to produce video content. The interface is built on SvelteKit (a JavaScript framework) and is connected to a Python backend server hosted on Google Cloud, which handles the generation and returns the results. For the generative models, we used Google's Gemini 2.5 Flash for a large language model, Veo 2 for a video generation model, and the Gemini 2.5 Flash Preview TTS model with the Zephyr prebuilt voice for a text-to-speech model. While these models were chosen as leading state-of-the-art options for their respective tasks, our system's modular design allows for the integration of alternative models into the pipeline.

5 Study

To understand the perspectives of researchers (PaperTok's target users) and the broader public (potential audiences of PaperTok videos), we conducted a two-part study consisting of a survey and a user study. The survey assessed perceptions of science communication videos generated by PDF-to-video platforms, while the user study focused on evaluating researchers' experience using PaperTok. Researcher participants completed both the survey and user study, while "broader public" participants (henceforth called "audience") participants only participated in the survey. We sought to address the following research questions:

RQ1: How do researchers and audiences perceive short-form science communication videos made by AI?

RQ2: In what ways do AI video tools support or hinder researchers' goals for communicating their work to broader audiences?

RQ3: What expectations and preferences do researchers have for the design of AI-enabled video tools for science communication?

RQ4: Are the videos generated using PaperTok perceived more positively (*e.g.*, credible and engaging) than those generated using available online PDF to short-form video services (*e.g.*, PDFtoBrainrot, SciSpace)?

5.1 Participants

5.1.1 Researchers. We recruited ($N = 18$) researchers through email outreach, university Slack channels, and online scholarly communities. Prior to the study, we verified that each was a named author in at least one published research paper and that they were willing to upload one of their papers to our AI system. Participants were compensated with $30 USD in an online gift card of their choice. These participants are aliased as R followed by a number.

Table 3: Summative study *researcher* participant demographics. (See Appendix B.1 for additional demographic details, and Appendix B.2 for *Audience* participant demographics.)

PID	Job title	PID	Job title
R1	Ph.D. student	R10	Ph.D. candidate
R2	Engineer, HCI researcher	R11	Ph.D. student
R3	Masters student	R12	Ph.D. student
R4	Ph.D. student	R13	Research fellow (Ph.D. in HCI)
R5	Research scientist	R14	Ph.D. student
R6	Ph.D. student	R15	Ph.D. student
R7	Ph.D. student	R16	Postdoctoral scholar
R8	Postdoctoral fellow	R17	Ph.D. candidate
R9	Ph.D. student	R18	Ph.D. student

5.1.2 Audience. We recruited ($N = 100$) participants using Prolific. We launched the study in 5 batches to cover different time zones and gather perspectives across the globe. Audience participants were compensated $10 USD via Prolific. These participants are aliased as A followed by a number.

5.2 Procedure

The following protocols were approved by our university's academic institutional review board (IRB).

5.2.1 Survey. We designed a two-part Qualtrics survey to gather participants' feedback on: (1) three AI-generated videos, (2) what they find acceptable or unacceptable use of AI in science communication, and (3) how credible and trustworthy they find AI-generated science communication videos. Similar to our formative study (see 3.2.1), each participant watched a set of three videos based on the same published CHI 2025 paper, but generated by a different AI tool: PDFtoBrainrot, SciSpace, and PaperTok. We chose the two comparators as they were two of the few available PDF to short-form video tools available. While both these platforms worked by simply having users upload a video, they represented two different approaches. PDFtoBrainrot, as the name suggests, overlays a voiceover summary of the paper using "brainrot language" on top of clips of video game play—representing a low-cost way of integrating scientific content with unrelated yet potentially engaging visuals to make the content digestible. SciSpace, on the other hand, provides a summary of the PDF with relevant images from the paper—focusing on creating video summaries of the paper. These comparators allow us to compare PaperTok videos against the fun and engaging videos and the more scientific video summaries.

Both participant types answered the same set of questions, except researchers selected which paper they wanted to watch videos of, while audience participants were randomly sorted into one of three sets. To prevent an ordering effect, the videos and questions were randomized in order for each participant.

Video evaluation measures. In the first part of the survey, participants evaluated each video on eleven dimensions. We measured (1) *engagement*, (2) *entertainment value*, and (3) *informativeness* based on shared content creator and science communication success metrics we discovered in our formative study. We also measured typical social media user behaviors such as (4) *shareability* and (5) *intent to watch more* as additional success metrics. In addition, we measured (6) *believability*, (7) *accuracy*, (8) *trustworthiness*, (9) *bias*, and
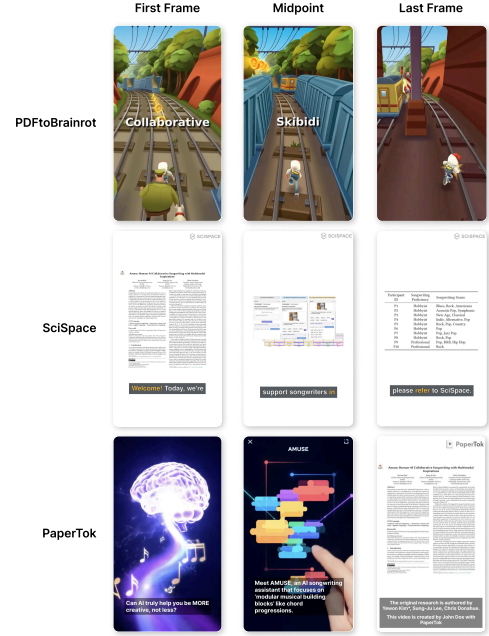

Figure 6: Screenshots from videos made using PDFtoBrainrot, SciSpace, and PaperTok based on the artifact contribution paper [25]

(10) *completeness* to assess information credibility [15]. Finally, we included a measure of (11) *production quality* to assess overall perceptions of the video's production value.

AI perception measures. In the second part of the survey, we measured the trustworthiness and acceptability of AI use in science communication, based on prior work that measured similar aspects but more broadly [19]. We also asked about perceptions of acceptability of AI use in specific aspects of science communication (*i.e.* doing background research, creating a script, and generating visual and audio content). For our complete set of questions and measures, see Appendix A.

5.2.2 PaperTok user study. We conducted a 60-minute user study via video call in pairs, with one study team member taking the lead on questions and the other taking notes on the participant's interactions with PaperTok. Each session began with the researcher participant taking the Qualtrics survey described in Section 5.2.1. During the survey, participants were provided privacy by turning off video and audio from all parties present.

After the survey, participants shared their screen and transformed one of their authored papers into a video, following the PaperTok user flow described in Section 4.1. During lull periods when PaperTok was processing (*e.g.*, turning script into a segmented storyboard), the interviewer asked targeted questions to gain insight into participant choices (*e.g.*, "Why did you choose that script over the others?"). At the end of the video-making process, participants viewed the resulting product and were asked to provide reactions.

On average, participants took about 20 minutes to generate their video. The remainder of the time was used for a semi-structured interview about their thoughts on their PaperTok video, overall experience using PaperTok, and perceptions of AI use in science communication.

5.3 Analyses

5.3.1 Quantitative results. To compare how participants evaluated the three different AI-generated videos (*i.e.*, PDFtoBrainrot, SciSpace, and PaperTok), we built mixed-effect models for the eleven Likert scale questions, using the Likert measure as the outcome variable, and video type, paper ID, and researcher or audience as fixed effects. We modeled participant ID as the random effect due to repeated measures. For the post-hoc analyses, we report on the results from Tukey's HSD for a more conservative approach to account for the multiple pair-wise comparisons (Figure 7).

Additionally, we conducted a one-way ANOVA to compare audience members ($N = 100$) and researchers ($N = 18$), exploring potential differences in their perceptions of trusting an AI tool for science communication (Figure 8).

Lastly, to understand our audience and researcher groups' acceptability of AI usage in science communication (categorical responses), we conducted a chi-squared test of independence on audience members and researchers' responses (Figure 9).

5.3.2 Qualitative results. We used a hybrid approach to thematic analysis: combining a codebook [5] with reflexive interpretation [6] to maintain methodological rigor while making space for the subjective experiences of participants and interviewers. To establish the initial codebook, four researchers independently coded the same three interviews, then met to compare interpretations and refine code definitions. After reaching consensus on the codebook, the remaining 15 interviews and 118 survey responses were divided among the four researchers for primary coding. Modifications to the codebook were encouraged and paired with asynchronous discussion in the team's Slack channel to ensure consensus while supporting the development of nuanced themes.

6 Results

In this section, we present the synthesized findings from the survey results and semi-structured interviews to describe how researchers and audience participants perceive AI in science communication.

6.1 Researchers' Perceptions of PaperTok

Creating engaging and informative science communication videos requires substantial time and effort. PaperTok serves as a prototype to this challenge, and researchers' generally positive reactions validate its intuitive interface and workflow. Furthermore, researcher reflections revealed key desirable features, like speedy workflow and clear attribution, and ways PaperTok can be used to foster science communication.

*6.1.1 **Functionality and value**.* The researcher participants' interactions with PaperTok and their interview responses revealed that the system can work as a low-effort way to create science communication videos.

Lowering barriers for research communication. All 18 participants rated PaperTok highly for its intuitiveness and efficient workflow. R5 remarked that anyone, *"even my mom can come in and use it without any problem."* Participants appreciated the stepped workflow and how *"with fairly minimal effort, [PaperTok] was able to capture a few of the main cruxes of the paper and its contribution...for maximum information summarization"* (R1).

In addition to speed, participants also highlighted the opportunity for creating *"more interesting visuals."* (R4) This was especially helpful for R14, who found it hard to create videos themselves because they *"don't have the expertise to know what visuals will be compelling... interesting to the general public."* They further described this challenge as being difficult to figure out if they would make a visual choice that made sense to them as the author, but would *"just overcomplicate it"* for *"someone who is just learning about [their research topic] for the first time."*

This difficulty in figuring out the right messaging and visuals is described by R4 as *"activation energy that tools like PaperTok could be helpful for."* R4 further added: *"I definitely recognize the value of having these short-form videos to disseminate content... especially to people who are probably not in academia. I definitely think that's important, but I just could not be bothered to do it myself without additional support."*

Use cases. When asked what PaperTok could best be used for, participants' answers revealed use cases beyond information-sharing, highlighting additional aspects of science communication tools that may be underdeveloped.

(1) *Share their research broadly.* Participants said they would use PaperTok as *"an engaging way to present to a general audience"* (R16). R13 further emphasized this point and noted it was *"easy to follow... compared to presenting for 20 to 40 minutes to expert people."* R12 believed PaperTok videos could make statistical results of studies easier to understand, which could in turn make these types of research more accessible and actionable compared to now. However, they cautioned that researchers using PaperTok should be *"very careful"* to minimize negative impact on the audience.

(2) *Generate ideas/support video prototyping.* Participants also mentioned that PaperTok was well-suited for generating ideas for phrasing and for visualizing their work, which they could then take and polish in other tools they have higher degrees of control with. This was especially true for R4, R17, and R18, who had experience making science communication videos. R14 felt this inspiration could be taken beyond video, stating, *"In my workflow right now, it would be most useful as kind of like an ideation thing [for] how am I going to distill my research... it might not end up in a video, [but] it might end up in a blog post, or something else."*

R18 described PaperTok as enabling rapid iteration by showcasing *"diverse"* visualization styles. They also noted how helpful it would be to post multiple versions of a video and get feedback from their target audience before working on a refined final product that synthesizes the most effective elements of the video prototypes.

(3) *Interest audience in their full paper.* Some participants felt a 45-second video is too short to fully communicate research, and instead saw it as a way to introduce broader audiences to their work and get them interested in reading the full paper. R8 recognized this challenge of getting the audience to read science, and felt that PaperTok can be helpful to remedy this.

(4) *Casually share with inner circle.* Some participants felt the current AI generation models are not *"there yet"* (R5) for broader science communication. However, they envision that it could be helpful to share PaperTok videos to others who are in *"the general STEM society"* (R11) to tell them about their research in an engaging way.

Credit scene as a credibility signal. A key feature of PaperTok is the credit scence (Figure 6), which automatically pulls the author list from the paper and generates an attribution overlay text that the user can further add to. Upon reaching this step, all 18 participants reacted positively. R1, R2, and R9 specifically mentioned this lends credibility to the video, with R9 commenting, *"It shows me that it's an actual paper, not just some random stuff on Twitter."* Researchers also liked how *"the contributions are made clear."* (R15) Some researchers suggested adding the publication venue (R11) or year (R9) to make this credit screen even better.

*6.1.2 **Thoughts on AI-Generated Assets.*** Participants thought the models did well in developing effective hooks and engaging scripts. However, when it came to the visual elements, participants raised concerns about the text-to-video generation model's inconsistent production quality and noted this as a potential blocker for fully utilizing PaperTok videos for science communication.

Script. PaperTok provided four sets of Hook and Body options for the participant to choose from. The majority of researchers ($N = 16$) were able to use a generated script without making any modifications. When asked why they chose the script, accuracy was most commonly cited as the deciding factor. R7 even chose the script that has a *"less exciting...big hook"* because *"all of the information is accurate in it, which has not been totally true for the [other] scripts."* For some researchers, this accuracy goes beyond getting facts of the paper right. For example, R5 chose a script that was also *"closer to how [they] would have sold this story, how [they] would have pitched this project."*

Half ($N = 9$) of the participants felt their research was oversimplified, but generally regarded it as a *"necessary"* (R9) by-product of condensing a lengthy research paper into a short video. R11 also noted that their paper contained highly technical jargon that they do not expect those outside of the field to know, so they *"wouldn't blame the video too much for [oversimplifying]."*

Visual. In the Storyboard step, participants were able to generate and regenerate visuals for their scenes. The participants found most of the generated clips to be acceptable, with some even exceeding expectations in being *"actually very realistic"* (R4) and depicting text correctly (R5). However, the visual style and quality proved to be inconsistent. Most participants encountered one or two visuals that had hallucinated details (*e.g.* *"glittery butterflies."* in R6's clip about mixed-abilities classrooms) or were irrelevant to their script

(*e.g.* chains showing up in R17's clip about potential stereotypes in AI models).

More complicated, even if the individual clips were acceptable, they often differed in style and mood from each other, which made the video feel *"sloppy"* (R10) once stitched together. Some participants noted that such discrepancies diminished the credibility of their work, making them unsuitable for professional presentation (R5, R11, R15).

Audio. The participants understood they could modify the text-to-speech model's prompt to affect tone and delivery of the voiceover, but almost all ($N = 15$) chose not to. After previewing the voiceover, participants were generally pleased and proceeded without edits. However, during the Storyboard step, some felt that the delivery and pacing of some clips could be improved, prompting R2 to wish for the same "trial-and-error" workflow for the audio generation, in conjunction with the visual and script iteration in the Storyboard step.

6.2 Perceptions of AI-Generated Videos

To evaluate how researchers and audiences perceived AI use in science communication and the resulting AI-generated videos, we analyzed their survey responses.

PaperTok videos were generally positively rated and had the highest mean ratings on eight of our eleven total dimensions. Informativeness ($M = 4.09$), believability ($M = 3.92$), and engagement ($M = 3.91$) were its' three highest-rated dimensions. Willingness to share, though, while still above neutral, had the lowest rating ($M = 3.05$, see Figure 7).

*6.2.1 **Video type comparisons**.* Across video platforms, we found that for the dimensions related to the informational quality of the videos, PDFtoBrainrot was consistently rated the lowest compared to SciSpace and PaperTok. This included all information credibility dimensions (believable, accurate, trustworthy, biased, and complete), and the informative rating. With our post-hoc HSD tests, we showed that PaperTok and SciSpace ratings were comparable (difference nonsignificant), and that they were both significantly higher than PDFtoBrainrot's ratings along these dimensions. This suggests that videos from PaperTok and SciSpace, but not PDFtoBrainrot, were positively perceived to provide informational value.

Comparing across other dimensions more related to engagement and entertainment, we observed that PaperTok's videos were rated significantly higher compared to the other two. PaperTok's videos were rated highest in terms of engaging ($M = 3.91$), entertaining ($M = 3.48$), and quality ($M = 3.71$). Participants also reported that they were more willing to watch more ($M = 3.50$) and more likely to share ($M = 3.05$) these videos. SciSpace's ratings were comparable to PDFtoBrainrot when it came to engagement ($M = 2.94$ vs $M = 2.92$, $p = 0.98$), but SciSpace was rated as less entertaining in comparison ($M = 2.44$ vs. $M = 2.81$, $p = 0.03$).

*6.2.2 **Audience vs. researcher perceptions**.* We analyzed researcher participants' and audience participants' differences in response to our question on the level of trusting AI for science communication (Appendix A). Our one-way ANOVA test (Figure 8)
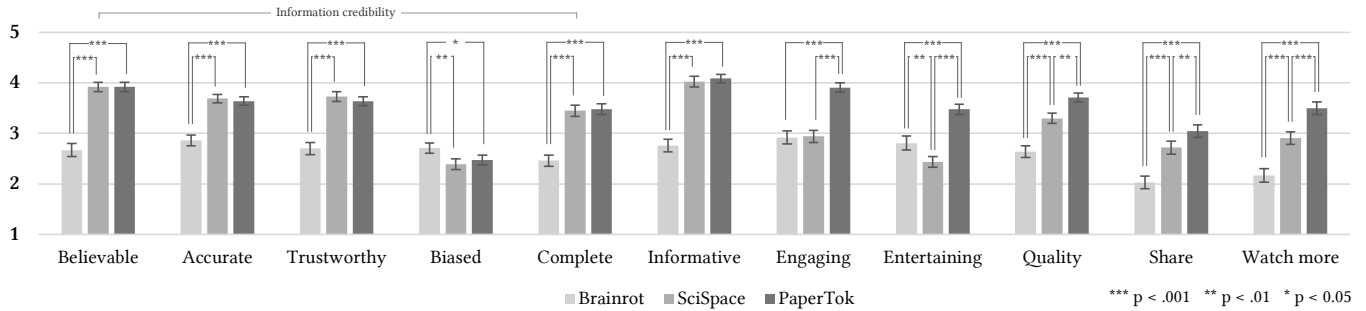
Figure 7: All participants' (N = 118) average ratings across 11 evaluation dimensions for each video type (PDFtoBrainrot, SciSpace, and PaperTok) on our 5-point Likert scale in vertical three-column bar chart format. Significance levels from our mixed-effects model estimated means analysis are denoted above between each video type. Bars indicate standard error.
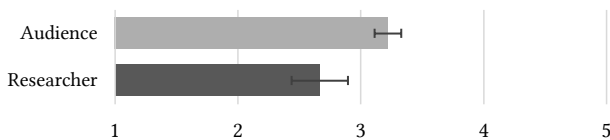

Figure 8: Audience members' and researchers' responses to *"On a scale of 1 to 5, how much do you agree with this statement: I trust the use of AI for science communication."* on our 5-point Likert scale in horizontal bar chart format. Audience members were significantly more trustworthy of AI use in science communication ($p < 0.05$) compared to researchers. Bars indicate standard error.
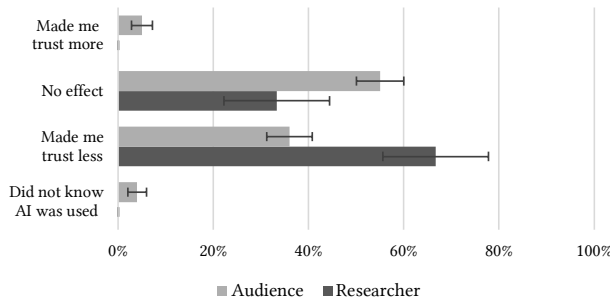

Figure 9: Audience members' and researchers' responses to *"Did knowing the videos you watched were created using AI affect your opinion of them?"* on categorical responses in horizontal bar chart format. When informed that videos were AI-generated, researchers showed a higher rate of decreased trust (66.7%) compared to audience members (36.0%). Audience members were more likely to respond with *"No effect"* (55.0%) compared to researchers (33.3%). Bars indicate standard error.

revealed that audience members were significantly more trustworthy of AI usage in science communication compared to researcher participants ($M = 3.22$ vs. $M = 2.67$, $F(1, 116) = 4.01$, $p < 0.05$).

We also explored whether researcher participants and audience participants differed in their evaluations of the acceptability of AI for science communication: did knowing that the videos were AI-generated affect perceptions between the two groups? (Figure 9). Our chi-squared analysis shows that the difference between the two groups is not highly significant but trending towards significance ($\chi^2 = 6.532$, $p < 0.1$). Audience participants less frequently reported decreased trust (36/100) compared to researchers (12/18), and were more likely to report no effect (55/100 vs. 6/18) or that it increased trust (5/100 vs. none). Four audience members indicated that they did not know AI was used.

6.3 Key Themes

In this section, we discuss three key themes that emerge from synthesizing both the researcher and audience participants' perspectives on the role of humans in AI-facilitated science communication.

6.3.1 Researchers desire a high level of precision in research communication. Our findings suggest that researchers desire fine-grained control over their science communication videos. As R16 puts it: *"I care a lot about this [research], so it's important to me that the findings [in the video] are accurate."* However, achieving this level of control is challenging due to limitations in text-to-video generation models and the nondeterministic quality of their output. This led to researchers feeling confused and frustrated when they could not determine why their script produced an undesirable visual detail. This frustration was especially difficult when they were satisfied with the script and wanted to adjust just the accompanying clip, but had no way to do it (R17). Other researchers were frustrated by common AI artifacts, such as *"random, blurry text"* (R5) and *"uncanny people"* (R2), that made an otherwise acceptable video unacceptable. Such *"sloppy mistakes"* (R6) were considered distracting errors that undermined their work. Audience participants shared this concern, with A78 noting, *"The video was good up to the point in which the vertical layout was no longer appropriate."* [5]

PaperTok attempted to provide this level of control by stepping users through scripting, voiceover selection, and storyboarding, but researchers wanted the option of an *"expert mode"* (R2) that allowed them to prompt the AI models directly. However, researchers also recognized that prompting itself could also be a challenge:

---

[5]See Figure 10 in Appendix C for a screenshot of the referenced clip.

"It feels difficult to just tune the model [or] the prompt to generate something I actually expect" (R2). To address existing limitations, participants proposed a range of mechanisms for increasing precision and control. Some suggested ways to shape overall style and communication strategy by providing a moodboard and customizing the voiceover audio's intonation. Others, like R13, suggested that PaperTok can incorporate figures from a paper directly, like what SciSpace does in their videos, to be *"more accurate in terms of information."*

However, audience responses suggest that figures from papers should be used sparingly and carefully. While some liked the *"the facts and data shown on screen to support the statements being made"* (A53), others disliked it for being *"hard to follow because it just showed the research text as the video"* (A19). Even worse, A2 perceived the primarily-figures SciSpace video as *"less credible, as the video is always showing image/text too small to read, making the visual part useless."* A25 agreed, feeling that they were shown *"random charts that might or might not be credible."* From the audience's perspective, what was most critical was using these figures effectively. Even though PaperTok does not currently incorporate these figures, A4 reported: *"What made it more credible was the fact that it explained thoroughly what the actual topic was about, using graphics and videos that related directly."*

6.3.2 Researchers and audiences differ in concerns about AI videos.
Researchers often worried that audiences would dismiss their videos if they appeared *"too AI-ish"* (R2). R16 even said they would be *"ashamed to share [their PaperTok] video with [their] participants"* because it continued to include *"nonsense text"* in the clips, despite multiple regenerations. Subtler artifacts, such as audio that sounded *"unnatural, as if it didn't know what punctuation was"* (R7) or had too much *"up-speak"* (R6), were also treated as threats to their credibility.

However, audience responses were more forgiving than what researchers expected. Rather than a blanket rejection, audience trust was more situational. Instead of nitpicking at AI artifacts, they judged primarily on production quality and effort. To them, the video's credibility hinged on whether the content was easy to understand (A4, A5, A10, A26, A28, A37), engaging (A7, A42, A64, A80), or polished (A12, A57, A60, A65). In other words, audiences were not as concerned about AI use affecting the credibility of the work, unless obvious mistakes drew their attention to it (A18, A25, A97, A98).

This gap between researcher concerns and audience reactions is also reflected in their differing acceptability ratings of AI usage in the science communication process. Researchers may have a heightened sensitivity to AI use in this context because they perceive these videos as permanent and public records of their scholarship. In R6's case, the AI artifacts in their PaperTok video made it *"sloppy"* and *"not reflective of [their] work,"*, which in turn made them reluctant to share it publicly.

These credibility concerns also extended to how authorship was represented in the credit scene. While researchers liked the attribution as paper authors, some felt uneasy being listed in a "Created by [name] with PaperTok" by-line. For them, "created" overstated their role in the process, because AI models generated the script, visuals, and audio. R5 described themselves as *"more like an approver,"*

and, like R7 and R13, preferred alternatives such as "Compiled by." R6 even removed their name entirely—leaving only "Created by PaperTok"—because they *"don't want to take credit because the AI did it."* Others, like R14, saw this by-line as *"crucial"* context for the production quality of the video. Together, these reactions highlight how questions of authorship and ownership further compound researchers' concerns about credibility in AI-generated videos.

6.3.3 Researchers and audience expect experts 'in the loop' in AI-facilitated science communication.
Both researchers and audience participants agreed that AI could play a supportive role in science communication, but only if mechanisms are in place to ensure accuracy. Both groups acknowledged that LLMs excel in processing large amounts of information efficiently (R13, A81), but are prone to hallucination (R1, R2, R4, R7, R14, R15, A18, A25, A29, A35, A66) or *"distortion"* (R3, R11, A8, A89). Therefore, signaling that the AI-generated video was based on rigorous scientific work and created with a human is a must. PaperTok accomplished this by having a human researcher go through each step with the AI tool, in addition to the credit scene that both participant groups appreciated as a credibility signal. However, some audience participants had higher expectations and expressed that *experts*—not just any human—should be in the loop in AI-facilitated science communication (A49, A57, A98).

Researchers have mixed opinions on their responsibility in this process. While R6 felt using AI-generated videos for science communication *"seems like taking away from science illustrators and human science communicators,"* R4 noted that, *"AI use shouldn't be vilified,"* with the caveat that researchers are *"double-checking the output."* R17 furthered this point, stating that *"AI is never reliable enough to accurately capture 100% of the nuance needed in science communication"* so researchers must be careful to use AI only in supportive roles (R3) and to *"make scientific findings"* (A20) by themselves. This shared view of using AI as an assistive tool rather than the main progenitor of ideas could be why both researchers and audiences liked the credit scene in PaperTok videos and considered it a boost to the credibility (R1-18, A27, A38, A39, A46).

With regards to *others* using PaperTok on their papers without their consent, participants generally viewed this as consistent with current practices—similar to journalists, bloggers, or other communicators summarizing their work. As R5 noted, *"Once the research is published, it's public knowledge."* However, as R1 and R3 requested, it would be better if PaperTok could provide a notification when their work was used in PaperTok, so they can check for inaccuracies and reach out to the creator to address them.

7 Discussion & Future Work

Advances in generative AI create a unique opportunity to support research communication by helping generate communication artifacts that may be too difficult or costly to create otherwise. In this work, we demonstrated that, with effective prompting and pipeline design, off-the-shelf LLMs can be leveraged to support the creation of short-form videos for research communication. Our AI-powered tool, PaperTok, was able to help researchers develop an engaging script and voiceover, create visually interesting scenes, and combine everything into a shareable video with a citation. Compared to videos generated by existing PDF-to-video platforms,

PaperTok-created videos were rated highly for being informative *and* engaging. Furthermore, although prior work has found that video-based delivery may struggle to capture different types of contributions [49], we demonstrated that PaperTok can generate videos for papers making artifact, empirical, and methodological contributions.

Our findings resonate with prior work that found scientists require precision in what is communicated [36], and that HCI researchers are concerned with inaccuracies, omissions, sensationalization, and the use of inappropriate tone for their work [42]. Importantly, our work revealed a key challenge in providing researchers with more nuanced control over AI-generated content. This is particularly needed for video, as, unlike text, there is no simple way to add, remove, or modify specific elements of the video without affecting the rest of it. While direct text-to-video model prompting may help, researcher participants expressed that fine-tuning prompts is difficult and that current interfaces do not support such fine-grained control. This highlights a key challenge for future systems to address in order to fully realize the potential of generative AI for research communication.

Using embedded visual content from the paper as video assets might address this challenge. However, since these figures and images are not specifically designed for video, they could be perceived as less engaging or lacking effort, which our findings indicate can diminish audience trust. Until text-to-video models or our ability to have fine-grained control improve, future iterations of PaperTok-like systems should at least provide users with the option to insert their own images or video clips that they have designed specifically for the video (*e.g.*, higher resolution images, system interaction clips).

A potential concern with lowering the barrier to creating short-form research communication videos is the risk of misuse or the spread of misinformation. Our design sought to address this in two key ways in the concluding credit scene—a valuable credibility signal that is often missing from AI-generated content. First, we included a screenshot of the referenced academic paper. This gives viewers indication that the source is peer-reviewed, while providing sufficient information to verify the source. Second, we encouraged PaperTok users to "sign-off" on their videos by including their names and credentials. These elements reflect a broader design implication that strong credibility signals enhance the trustworthiness of the content, by making authorship and accountability visible. In essence, our approach relies on the video creator to verify the accuracy of the information and stake their own credibility in the process.

Furthermore, this type of watermarking provides the creator acknowledgment for their work and opportunity to gain recognition for it. This branding opportunity can serve as incentive for people to create videos while ensuring the content's accuracy, as their reputation is at stake. Both researchers and audience found this to be an effective design. To further enhance trust, researcher participants suggested adding a feature in PaperTok that notifies and invites authors of the papers to review the generated video and provide feedback on accuracy and messaging. Collectively, these design considerations point toward a model of AI-assisted science communication, where responsibility is distributed but traceable, and where human and AI contributions are purposefully distinguished rather than blurred.

Ultimately, our work points to longer-term questions that need to be studied. Our work only explored the creation of videos and did not evaluate their impact "in-the-wild." Future work should explore ways to improve its reach or virality to help broaden the dissemination of these videos. Consequently, researchers should also critically examine how this use of generative AI may transform science communication. With tools like these (and with the already exponential growth in published papers [4]), there could be a large influx of research communication content shared online. On one hand, it could replace "brain rot" content—which could lead to the deterioration of a person's mental or intellectual state [51]—and improve social media experiences. On the other hand, short-form video content could also negatively affect and replace people's deeper engagement with the underlying science. As short-form AI-generated research videos become more widespread, understanding how to balance information accessibility, trust, and meaningful engagement will be essential for the future of science communication.

8 Limitation

While our system is adaptable to research papers across diverse domains, our evaluation concentrated on HCI papers as a specific form of research communication. Our user study participants were also primarily Ph.D. students and early-career researchers, as they represent our target users of PaperTok. Future work should investigate whether patterned differences in use or design implications emerge when applied to other domains or participant groups. In addition, our system offered only text-to-speech voiceovers for audio options. While users could modify the prompt to adjust tone, other personalization factors, such as music or audio effects, were not provided for. This constrained participants' ability to fully personalize the delivery of their findings through the video, which could be a factor in lower levels of perceived ownership over the final product.

9 Conclusion

We introduce PaperTok as an AI support tool for researchers to create accessible science communication for broader audiences. We designed a workflow that centralized the human in key decision points, while AI handled the early, labor-intensive steps of condensing scholarly papers into a punchy, 45-second script and generating matching visuals for it. Researchers acted as creative directors as they chose, refined, and shaped their short-form video scene by scene to suit their voice and expertise.

Our studies indicate this approach *can* work: science communication can be engaging for audiences and less costly for researchers. However, it also surfaced a key tension between researcher and AI when participants found value in the efficiency of the workflow but felt limited ownership when they could not precisely control visuals due to limitations in text-to-video generation. These findings suggest that the real opportunity of generative AI lies not in replacing the totality of the creative labor in creating science communication, but in enabling richer forms of human-in-the-loop collaboration with AI.

Ultimately, we see our paper contributing to the ongoing conversation about human-AI co-creation, what it means to responsibly integrate AI support tools into science communication, and how we might design systems to help people communicate and connect with research.

References

[1] Amanda Alvarez, Aylin Caliskan, MJ Crockett, Shirley S Ho, Lisa Messeri, and Jevin West. 2024. Science communication with generative AI. *Nature Human Behaviour* 8, 4 (2024), 625–627.

[2] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2023. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* 30, 5 (2023), 1–38.

[3] Sibusiso Biyela, Kanta Dihal, Katy Ilonka Gero, Daphne Ippolito, Filippo Menczer, Mike S Schäfer, and Hiromi M Yokoyama. 2024. Generative AI and science communication in the physical sciences. *Nature Reviews Physics* 6, 3 (2024), 162–165.

[4] Lutz Bornmann, Robin Haunschild, and Rüdiger Mutz. 2021. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–15.

[5] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. Sage.

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[7] Anna R. L. Carter, Dalya Al-Shahrabi, William A. P. Imoukhuede, Lauren Scott, Dan Jackson, Reem Talhouk, Patricia Piedade, and Kyle Montague. 2025. KaraokAI: Exploring the Potential of AI-Generated Karaoke for Academic Conferences and Research Experiences. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 734, 5 pages.

[8] Isabelle M Côté and Emily S Darling. 2018. Scientists on Twitter: Preaching to the choir or singing from the rooftops? *Facets* 3, 1 (2018), 682–694.

[9] Emily S Darling, David Shiffman, Isabelle M Côté, and Joshua A Drew. 2013. The role of Twitter in the life cycle of a scientific publication. *arXiv preprint arXiv:1305.0435* (2013).

[10] Sarah R Davies. 2008. Constructing communication: Talking to scientists about talking to the public. *Science communication* 29, 4 (2008), 413–434.

[11] Luigi De Angelis, Francesco Baglivo, Guglielmo Arzilli, Gaetano Pierpaolo Privitera, Paolo Ferragina, Alberto Eugenio Tozzi, and Caterina Rizzo. 2023. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in public health* 11 (2023), 1166120.

[12] Elaine Howard Ecklund, Sarah A James, and Anne E Lincoln. 2012. How academic biologists and physicists view science outreach. *PloS one* 7, 5 (2012), e36240.

[13] Maha Eldamnhory. 2024. *Academify-Lite: LLM-powered social media post generator for academic content*. Master's thesis. M. Eldamnhory.

[14] Wiebke Finkler and Bienvenido León-Anguiano. 2019. The power of storytelling and video: a visual rhetoric for science communication. (2019).

[15] Andrew J Flanagin and Miriam J Metzger. 2000. Perceptions of Internet information credibility. *Journalism & mass communication quarterly* 77, 3 (2000), 515–540.

[16] Naomi Forman-Katz and Katerina Eva Matsa. 2022. News platform fact sheet. *Pew Research Center* 20 (2022). https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/

[17] Claudia Frick, Eva-Maria Grommes, and Alice Watanabe. 2025. Science communication on TikTok: toward transformative and post-normal science. *Frontiers in Communication* 10 (2025), 1491250.

[18] Anna Gardiner, Miriam Sullivan, and Ann Grand. 2018. Who are you writing for? Differences in response to blog design between scientists and nonscientists. *Science Communication* 40, 1 (2018), 109–123.

[19] Michael Gerlich. 2023. Perceptions and Acceptance of Artificial Intelligence: A Multi-Dimensional Study. *Social Sciences* 12 (2023).

[20] Silas Hsu, Vinay Koshy, Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2025. Placebo Effect of Control Settings in Feeds Are Not Always Strong. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

[21] Philip Hunter. 2016. The communications gap between scientists and public: More scientists and their institutions feel a need to communicate the results and nature of research with the public. *EMBO reports* 17, 11 (2016), 1513–1515.

[22] Hepeng Jia, Dapeng Wang, Weishan Miao, and Hongjun Zhu. 2017. Encountered but not engaged: Examining the use of social media for science communication by Chinese scientists. *Science Communication* 39, 5 (2017), 646–672.

[23] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hsieh. 2019. r/science: Challenges and opportunities in online science communication. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[24] Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. 2025. 'I'm Categorizing LLM as a Productivity Tool': Examining Ethics of LLM Use in HCI Research Practices. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–26.

[25] Yewon Kim, Sung-Ju Lee, and Chris Donahue. 2025. Amuse: Human-AI Collaborative Songwriting with Multimodal Inspirations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–28.

[26] Clayton T Lamb, Sophie L Gilbert, and Adam T Ford. 2018. Tweet success? Scientific communication correlates with increased citations in Ecology and Conservation. *PeerJ* 6 (2018), e4564.

[27] Gionnieve Lim and Simon T Perrault. 2024. Evaluation of an llm in identifying logical fallacies: A call for rigor when adopting llms in hci research. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. 303–308.

[28] Xinru Lin and Luyang Li. 2025. Implicit bias in llms: A survey. *arXiv preprint arXiv:2503.02776* (2025).

[29] María José Luzón. 2013. Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written Communication* 30, 4 (2013), 428–457.

[30] Uday Mittal, Siva Sai, Vinay Chamola, et al. 2024. A comprehensive review on generative AI for education. *IEEE Access* (2024).

[31] Meredith Ringel Morris. 2023. Scientists' Perspectives on the Potential for Generative AI in their Fields. *arXiv preprint arXiv:2304.01420* (2023).

[32] Matthew C Nisbet and Dietram A Scheufele. 2009. What's next for science communication? Promising directions and lingering distractions. *American journal of botany* 96, 10 (2009), 1767–1778.

[33] Obinna Ogbonnia Otuu and Deepak Sahoo. 2025. How Should We Design Technology With Diverse Stakeholders Who Wish Not to Attend Design Activities Together?. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–14.

[34] Will Douglas Heavenarchive page. [n. d.]. Why Meta's latest large language model survived only three days online. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/

[35] Hans Peter Peters. 2013. Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences* 110, supplement_3 (2013), 14102–14109.

[36] Andrew M Petzold, Robert L Dunbar, and Marcia D Nichols. 2025. Bridging the gap: Understanding why scientists find it difficult to communicate their ideas. In *Truth Unveiled*. Elsevier, 181–199.

[37] Ben Rein. 2023. Making science education more accessible: A case study of TikTok's utility as a science communication tool. *Neuroscience* 530 (2023), 192–200.

[38] Mike S Schäfer. 2023. The Notorious GPT: science communication in the age of artificial intelligence. *JCOM: Journal of Science Communication* 22, 2 (2023), Y02.

[39] Dietram A Scheufele. 2013. Communicating science in social settings. *Proceedings of the National Academy of Sciences* 110, supplement_3 (2013), 14040–14047.

[40] Yi Shi and Lin Sun. 2024. How generative AI is transforming journalism: Development, application and ethics. *Journalism and Media* 5, 2 (2024), 582–594.

[41] Donghoon Shin, Lucy Lu Wang, and Gary Hsieh. 2024. From paper to card: transforming design implications with generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[42] C Estelle Smith, Eduardo Nevarez, and Haiyi Zhu. 2020. Disseminating research news in HCI: Perceived hazards, how-to's, and opportunities for innovation. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[43] C Estelle Smith, Xinyi Wang, Raghav Pavan Karumur, and Haiyi Zhu. 2018. [Un]breaking News: Design Opportunities for Enhancing Collaboration in Scientific Media Production. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[44] Mohamad-Hani Temsah, Rakan Nazer, Ibraheem Altamimi, Raniah Aldekhyyel, Amr Jamal, Mohammad Almansour, Fadi Aljamaan, Khalid Alhasan, Abdulkarim A Temsah, Ayman Al-Eyadhy, et al. 2025. OpenAI's sora and google's veo 2 in action: a narrative review of artificial intelligence-driven video generation models transforming healthcare. *Cureus* 17, 1 (2025).

[45] Luxuan Wang, Galen Stocking, Samuel Bestvater, and Regina Widjaya. 2025. A closer look at Americans' experiences with news on TikTok. *Pew Research Center* (2025). https://www.pewresearch.org/short-reads/2025/01/17/a-closer-look-at-americans-experiences-with-news-on-tiktok/

[46] Pengqi Wang, Mengli Yu, and Yadong Liu. 2022. Assessing the Content Topics of the Educational Videos on Tik Tok for Science Communication. In *2022 6th International Seminar on Education, Management and Social Sciences (ISEMSS 2022)*. Atlantis Press, 1792–1801.

[47] Sitong Wang, Samia Menon, Tao Long, Keren Henderson, Dingzeyu Li, Kevin Crowston, Mark Hansen, Jeffrey V Nickerson, and Lydia B Chilton. 2024. ReelFramer: Human-AI co-creation for news-to-video translation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[48] Dustin J Welbourne and Will J Grant. 2016. Science communication on YouTube: Factors that affect channel and video popularity. *Public understanding of science* 25, 6 (2016), 706–718.

[49] Spencer Williams, Ridley Jones, Katharina Reinecke, and Gary Hsieh. 2022. An HCI research agenda for online science communication. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.

[50] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44.

[51] Ahmed Mohamed Fahmy Yousef, Alsaeed Alshamy, Ahmed Tlili, and Ahmed Hosny Saleh Metwally. 2025. Demystifying the new dilemma of brain rot in the digital era: A review. *Brain Sciences* 15, 3 (2025), 283.

[52] Shupei Yuan, Shaheen Kanthawala, and Tanya Ott-Fulmore. 2022. "Listening" to science: Science podcasters' view and practice in strategic science communication. *Science Communication* 44, 2 (2022), 200–222.

[53] Emily E Zawacki, Wendy Bohon, Scott Johnson, and Donna J Charlevoix. 2022. Exploring TikTok as a promising platform for geoscience communication. *Geoscience Communication* 5, 4 (2022), 363–380.

A Survey Questions
A.1 Video Feedback

The following questions in each block were randomized for each participant, except for the final question on overall production quality and final open-ended question which were fixed at the end of the survey. All questions were on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree), except for production quality, whose scale was from 1 (very poor quality) to 5 (very good quality).

A.1.1 Block 1.

- The video was engaging.
- The video was entertaining.
- I learned something from the video.
- I would share this video with others.
- I would watch more videos like this from the same creator.
- Rate the overall production quality of the video.

A.1.2 Block 2.

- The information presented in the video was believable.
- The information presented in the video was accurate.
- The information presented in the video was trustworthy.
- The information presented in the video was biased.
- The information presented in the video was complete.
- What made the video seem more or less credible?

A.2 AI Perceptions

The following questions were based on a 5-point Likert scales were from 1 (completely disagree) to 5 (completely agree) for the trust question, and 1 (highly unacceptable) to 5 (highly acceptable) for the acceptability questions.

- Did knowing the videos you watched were created using AI affect your opinion of them?
  - Yes, it made me trust in the video content more
  - Yes, it made me trust the video content less
  - No, it didn't affect my opinion
  - I didn't know AI was used until now
- On a scale of 1 to 5, how much do you agree with this statement: I trust the use of AI for science communication.
- Why do or don't you trust the use of AI for science communication?

- On a scale of 1 to 5, how acceptable do you find the use of AI tools in each of these aspects?
  - Do background research?
  - Create a script for a science communication video?
  - Generate visual content (images and video) for science communication?
  - Generate audio content (images and video) for science communication?

A.3 Demographics

- How often do you use social media for learning or gaining information?
- How often do you watch short-form videos (*e.g.*, Tiktok, YouTube Shorts, Instagram Reels)?
- How often do you read content about science or academic research?
- What is your age?
- What is your gender?
- What is your highest level of education?

B Summative Study Participants

B.1 Researcher Participant Demographics

Table 4: Researcher participant demographics ($N = 18$)

Category	N (%)
Age	
18-24	3 (16.7%)
25-34	15 (83.3%)
Gender	
Female	9 (50.0%)
Male	7 (38.9%)
Non-binary	1 (5.6%)
Prefer to self-describe	1 (5.6%)
Education	
Bachelor's degree	5 (27.8%)
Master's degree	6 (33.3%)
Doctoral degree	7 (38.9%)
Behavior	**N (%)**
Social Media for Learning	
1-2 times per week	2 (11.1%)
3-4 times per week	1 (5.6%)
5-6 times per week	3 (16.7%)
Once daily	2 (11.1%)
2-5 times daily	6 (33.3%)
6-9 times daily	1 (5.6%)
10-13 times daily	2 (11.1%)
Hourly or more	1 (5.6%)
Short-form Video Consumption	
3-4 times per week	1 (5.6%)
5-6 times per week	1 (5.6%)
Once daily	4 (22.2%)
2-5 times daily	6 (33.3%)
6-9 times daily	1 (5.6%)
10-13 times daily	4 (22.2%)
Hourly or more	1 (5.6%)
Science Content Reading	
2-3 times a week	2 (11.1%)
Almost every day	8 (44.4%)
Every day or more	8 (44.4%)

B.2 Audience Participant Demographics

Table 5: Audience participant demographics ($N = 100$)

Category	N (%)
Age	
18-24	16 (16.0%)
25-34	34 (34.0%)
35-44	23 (23.0%)
45-54	15 (15.0%)
55-64	10 (10.0%)
65 and above	2 (2.0%)
Gender	
Female	49 (49.0%)
Male	51 (51.0%)
Education	
High school or equivalent	9 (9.0%)
Some college/trade school	20 (20.0%)
Bachelor's degree	48 (48.0%)
Master's degree	15 (15.0%)
Doctoral/Professional degree	7 (7.0%)
Prefer not to say	1 (1.0%)
Behavior	**N (%)**
Social Media for Learning	
Never	4 (4.0%)
1-2 times per week	7 (7.0%)
3-4 times per week	10 (10.0%)
5-6 times per week	8 (8.0%)
Once daily	16 (16.0%)
2-5 times daily	24 (24.0%)
6-9 times daily	13 (13.0%)
10-13 times daily	8 (8.0%)
Hourly or more	10 (10.0%)
Short-form Video Consumption	
Never	1 (1.0%)
1-2 times per week	9 (9.0%)
3-4 times per week	7 (7.0%)
5-6 times per week	5 (5.0%)
Once daily	7 (7.0%)
2-5 times daily	24 (24.0%)
6-9 times daily	8 (8.0%)
10-13 times daily	23 (23.0%)
Hourly or more	16 (16.0%)
Science Content Reading	
Never	2 (2.0%)
Less than once a month	9 (9.0%)
About once a month	5 (5.0%)
2-3 times a month	18 (18.0%)
About once a week	13 (13.0%)
2-3 times a week	25 (25.0%)
Almost every day	21 (21.0%)
Every day or more	7 (7.0%)

C   PaperTok Video Sample

Figure 10: Screenshot of the clip referenced in Section 6.3.1