LUND UNIVERSITY

AI and authenticity

Young people's practices of information credibility assessment of AI-generated video content

Lao, Yucong; Hirvonen, Noora; Larsson, Stefan

Published in:
Journal of Information Science

2025

Document Version:
Publisher's PDF, also known as Version of record

[Link to publication](#)

Total number of authors:
3

AI and authenticity: Young people's practices of information credibility assessment of AI-generated video content

Yucong Lao
Faculty of Humanities, University of Oulu, Finland

Noora Hirvonen
Faculty of Humanities, University of Oulu, Finland

Stefan Larsson
Department of Technology and Society, Lund University, Sweden

Abstract

The breakthroughs in generative artificial intelligence (AI) technologies have enabled the creation of hyper-realistic media content that credibly mimics human beings. As AI-generated media (AGM) become more accessible on social media platforms, concerns about young people's capabilities to deal with mis- and disinformation are rising. This study aims to investigate young people's practices to assess the credibility of AGM. This study applies interviews with an experiment offering two selected videos to investigate young people's strategies for verifying the authenticity of AI-generated deepfake videos. In addition, it examines their reflections on the implications of such media. The results indicate how young people verify the authenticity of AGM based on image, sound, narrative, and intuition and how they sense mis- and disinformation from three main perspectives: techniques, consequences, and associated people. This study contributes to the exploration of AGM from the perspective of young people's media and information literacy practices.

Keywords

AI-generated media; authenticity verification; deepfake; information credibility evaluation; media and information literacy; young people

1. Introduction

1.1. Background

In the digital era, artificial intelligence (AI) is increasingly being used to produce media in different forms. The term AI-generated media (AGM) is used to refer to content created by an AI system in the shape of audio, video, image, text and other modalities [1], which can be utilised in various scenarios. For example, at the Dalí Museum in St Petersburg, Florida, a surrealist painter used AGM tools to present life-size avatars of Dalí on digital screens [2], and they are used in the entertainment industry to enhance the quality of dubbing on foreign-language films [3]. In everyday language, these AI-powered artefacts are widely termed 'deepfake', 'deep fake' or 'synthetic media' [4–6]. According to Vaccari and Chadwick [6], deepfakes refer to videos in which someone's facial expressions are pasted onto the head of another, powered by machine learning algorithms. While AGM has a longer history than the often-cited breakthrough for a particular methodology of generative adversarial network (GAN) made in 2014 [7,8], recent advancements of so-called

Corresponding author:
Yucong Lao, University of Oulu 90570, Pentti Kaiteran Katu 1, Oulu 90570, Finland.
Email: yucong.lao@oulu.fi

large language models (LLMs) have led to a range of remarkably capable AGM tools like Dall-E 3[1], ChatGPT[2] and multimodal models which further stress the imminent need to understand their implications.

While AGM offers unprecedented possibilities for creativity and innovation, it also presents great challenges for society. The significant breakthroughs in generative AI technologies have enabled the production of hyper-realistic content whose authenticity is difficult for audiences to verify [9–11]. For example, deepfake videos, especially those in low resolution, are 'alarmingly convincing' and, as such, contribute to spreading disinformation [6]. With the rapid increase of AGM [12], they can be used to propagate disinformation and enable fraudulent activities, posing a high risk to individuals, organisations, and society [13].

As AGM spread ubiquitously on social media platforms, they have become very accessible. Hence, concerns have been expressed about young people's vulnerability to such content. As Ali et al. [14] stress, AGM is widely circulating on social media frequented by young people [15], who might not be aware of the existence of this content. Moreover, both news articles and research indicate that young people are probably to believe in and even spread misinformation [16,17]. From the perspective of media and information literacy (MIL), AGMs are problematic because they challenge our accustomed practices to 'critically evaluate the information providers for authenticity, authority, credibility, and current purpose, and weighing up opportunities and potential risks' [18]. In this context, it is important to understand how young people deal with AGM content to identify potential benefits and counter risks and harms.

1.2. Purpose and research questions

The overall aim of the study is to contribute to the exploration of AGM from the perspective of young people's media and information literacy (MIL) practices. It does so by investigating young Finns' practices to assess the credibility of AGM content, focusing specifically on AI-generated deepfake videos. Under this aim, the following research questions are to be answered:

> *RQ1. How is the authenticity of the media content in AI-generated deepfake videos verified?*
> *RQ2. How are the implications of AI-generated deepfake videos understood and reflected on?*

The research was conducted in Finland, which is a country characterised by a high level of trust in public institutions [19–21], high media literacy [22], and high Internet penetration (around 97% according to several estimates). Moreover, Finland is considered devoted to promoting young people's media literacy [23]. These factors make Finland a representative context for examining MIL, as it offers a benchmark for comprehending young people's capabilities to critically assess and navigate AGM content.

The article is structured as follows: the next section reviews the related literature on MIL in the context of mis- and disinformation and information credibility assessment. Following this, the section of methods introduces the research design, the process of participant recruitment, data collection and analysis, and the ethical considerations. The subsequent section presents the findings, including young people's strategies of authenticity verification and their sense of mis- and disinformation in the context of AGM. Then, the study discusses these findings in the context of existing research and explores their broader implications. In conclusion, this research provides a summary of the key insights derived from the study.

2. Literature review

This section primarily includes two parts. First, it reviews the existing literature on MIL and emphasises its significance in an age saturated with mis- and disinformation. Second, it unfolds the previous research related to information credibility assessment from two perspectives – the definitions of credibility and the strategies of information credibility assessment; furthermore, authenticity verification is identified as an approach utilised in this study.

2.1. Situating MIL in the context of mis- and disinformation

MIL is a concept that merges both media literacy and information literacy [24,25]. UNESCO has been actively promoting both research and practical strategies on MIL to enhance and deepen the understanding of this concept and promote it among citizens. According to UNESCO [26,cited by 27], MIL is:

> [...] a set of competencies that empowers citizens to access, retrieve, evaluate and use, create as well as share information and media content in all formats, using various tools, in a critical, ethical and effective way, in order to participate and engage in personal, professional and societal activities.

In their publication, Grizzle et al. [28], in collaboration with UNESCO, expanded the concept of MIL. This expanded concept includes the knowledge, skills, and attitudes that empower citizens to critically evaluate content, ethically locate and use information, apply information and communication technology (ICT) skills, and engage in self-expression and democratic participation.

While advanced technologies have fuelled information and media creation and dissemination, they simultaneously contribute to the emergence and spread of mis- and disinformation. Confronting this situation, many scholars have particularly underlined the significance of *information evaluation* under the umbrella of MIL. For example, Haider and Sundin [29] noted that while facing an information crisis, we should stress the essentialness of 'gauging fact, trustworthiness and credibility', as a vital part of MIL. In addition, much literature addresses the perspective of critical thinking. In their conceptual article, scholars Brisola and Doyle [30] equated MIL with critical information literacy (CIL), asserting that critical and reflexive thinking is vital for individuals to effectively counteract mis- and disinformation. In line with Brisola's and Doyle's arguments [30], Abu-Fadil [31] also projected that guiding students to cultivate the skills of critical thinking is a path to be ready for combating mis- and disinformation, especially from media content, such as fake news.

A large share of existing research on MIL has focused on issues related to so-called fake news. For instance, Tandoc Jr et al. [32] conducted in-depth interviews with 20 Singaporean participants to explore their approaches to handling mis- and disinformation, indicating that a majority of participants exhibited indifference towards encountering fake news unless the issues highlighted by the misinformation strongly resonated with their personal relevance. By employing an experimental design on a sample of 100 Jordanian undergraduate students, Reem [33] argues that MIL can help students to analyse various materials and increase the possibility of identifying inauthentic content from news, photos, and videos.

As the information infrastructure of society evolves, our research focus on MIL should also shift accordingly [29]. With numerous tech companies' contributions to generative AI technology, such as OpenAI[3]'s development of ChatGPT and Sora[4], AGMs are on the rise, and therefore bring both benefits and challenges to the media and information landscape. While the widespread dissemination of indistinguishable information and media content generated by AI poses a considerable challenge, there has been a limited amount of research conducted on people's MIL regarding AGM. As an exception to this gap, Vaccari and Chadwick [6] examined 2005 British participants' responses to the three variants of the BuzzFeed Obama/Peele video[5] and found that people were more prone to experiencing uncertainty rather than being misled by deepfakes. However, given that the video provided by researchers was among the most widely recognised deepfakes at that time, some participants may have been aware of their falseness, potentially influencing the results. In addition, by investigating 2689 audience comments under the top 10 YouTube deepfake videos, Lee et al. [34] found that most of the comments expressed neutral or irrelevant attitudes towards these videos. Yet, it's noteworthy that the participants recruited by the researchers exclusively represented the adult demographic. In this context, as young people are especially vulnerable to being negatively impacted by the mis- and disinformation of AGM, the present study specifically addresses young people's MIL practices regarding deepfakes, to contribute to filling the gap existing in this research field.

Furthermore, while discussing MIL regarding ICT-related topics (e.g. digital tools, technologies such as AI, etc.), much of the existing research involves technology as one of the pivotal perspectives. For example, UNESCO's MIL Curriculum for Teachers [35] includes the elements of learning and using ICT skills for information processing and media content production in MIL. As advanced technology becomes increasingly entangled with the production and distribution of information, it is inevitable to consider technology's role in examining people's MIL. Therefore, as indicated by RQ1 and RQ2 above, this article considers how young people's MIL concerning AGM is shaped by their understanding of AI technologies.

2.2. Information credibility assessment

2.2.1. Defining credibility. Under the umbrella of MIL, and as indicated in the aim above, this study examines how young people evaluate the content of AGM by focusing on their credibility assessment practices. The discussion on the concept of credibility has been active in different fields, including human–computer interaction (HCI), psychology, and library and information science (LIS) [36]. For instance, Tseng and Fogg [37] emphasise the importance of people's thoughts and experiences on credibility in the context of HCI and distinguish between four types of credibility based on interactions between users and computer systems:

- Presumed credibility: Pertains to the degree of belief a perceiver holds due to general assumptions in their mind.
- Reputed credibility: Relates to the extent of belief based on reports from third parties.
- Surface credibility: Refers to the level of belief a perceiver has based on initial inspection.
- Experienced credibility: Refers to the degree of belief formed through direct, firsthand experience.

In the field of LIS, credibility is considered to be a complex concept that is often defined together with concepts such as 'believability, trustworthiness, fairness, accuracy, trustfulness, factuality, completeness, precision, freedom from bias, objectivity, depth and informativeness' [38]. Furthermore, Rieh [38] contends that, in a general sense, credibility is regarded as 'people's assessment of whether the information is trustworthy based on their own expertise and knowledge'.

Notably, scholars conceptualising credibility as people's perception and belief of certain objects they have received [39,40] have commonly utilised terms such as 'perceived credibility of [...]' [41] or 'perception of [...] credibility' [42]. From this perspective, individuals' characteristics, knowledge, and experienced norms can be viewed to be shaped by their social environment, which in turn influences their perceptions of information credibility and impacts their credibility judgements [43,44].

Scholars have categorised credibility into various types, primarily including source credibility, message credibility and media/medium credibility [38,45]. These categories have been extensively applied in numerous studies within LIS and media and communication research [38,46]. According to Rieh [38], the explanations can be summarised into three types of credibility: *source credibility* underlines an information receiver's judgements considering a communicator's believability; *message credibility* stresses how the features of a message, including content, structure, language, and presentation, influence information receiver's perception of information believability and *media credibility* highlights information receiver's judgements on the medium through which a message is sent.

According to Metzger et al.'s literature review [47] related to source credibility, much previous research [48,49] points out that source credibility is often concerned with two dimensions –'trustworthiness' (e.g. the perceived friendliness and attractiveness of a source can enhance its trustworthiness among people) and 'expertise' (e.g. information provided by highly qualified and reliable speakers is often regarded as highly credible). Source identification is vital to detecting credible content by verifying 'authors' expertise, affiliation or interests to judge the trustworthiness of information' [50].

However, verifying source credibility can be challenging. On one hand, information in an online environment is often 'presented seamlessly' with a blurred boundary between 'advertising and information' [42]. On the other hand, with advanced technology, both the information sender and the information itself can be indistinguishably fabricated, impacting people's judgements of trustworthiness and expertise. What's more, as people are overwhelmed by the amount and dissemination speed of information in their everyday lives, they do not have time or intention to check the source of each piece of information they encounter [30]. Consequently, the content of the message itself frequently becomes the primary basis for people to form their judgements. This underscores the increased relevance of message credibility assessment, as the message is the entity directly received and evaluated by users. Hence, in this study, we focus on message credibility as a perspective to assess the credibility of AGM.

2.2.2. Strategies of information credibility assessment.
People's strategies for assessing credibility have been examined from diverse theoretical and methodological perspectives. Widely applied models for analysing individuals' ways of information credibility evaluation include the heuristic–systematic model (HSM) and the elaboration likelihood model (ELM). HSM involves two types of processing –*systematic processing*, which requires the information receiver to possess both the fundamental ability and capacity to analyse information, and *heuristic processing*, which necessitates that the information receiver has basic judgement-relevant knowledge [51,52]. The ELM presents dual routes of how people deal with information, including 'central' and 'peripheral' routes. The central route focuses on making an effort to process information and aiming at solving the central issues revealed in the given information; the peripheral route costs less cognitive effort in information processing [53,54]. Notably, both models stress information receivers' motivation while confronting information in a specific context.

Beyond a model-based approach, a large body of research has focused on either a specific group of people's information credibility assessment or people's credibility assessment of different types of information. From the former perspective, much of the existing research is about young people's strategies for information credibility assessment [55]. For example, Subramaniam et al. [55] employed an ethnographic approach to explore disadvantaged tweens' (aged 11–13 years) ways of assessing the credibility of online information and found the factors that impact their credibility assessment strategies, such as limited English vocabulary, unfamiliarity with otherwise well-known sources and preference for non-textual formats like audio and video. Moreover, Kohnen et al. [56] investigated eighth-grade students' strategies for credibility assessment of unfamiliar websites, indicating their predominant use of strategies associated with 'reading of

informational texts' and 'evaluating websites'. Notably, albeit using these strategies, participants were unable to offer warranted explanations for their credibility judgements, nor were they capable of identifying biased websites.

Concerning the latter perspective, numerous studies have concentrated on how people evaluate online information presented by various media/mediums [57]. For example, Lederman et al. [57] examined users' ways of assessing the credibility of information in a health forum, identifying five criteria used by participants, including 'reference credibility', 'argument quality', 'verification', 'contributor's literacy competence' and 'crowd consensus'. Furthermore, by employing a systematic literature review of over 100 papers, Shah et al. [58] pointed out that combining techniques that measure multiple perspectives, including the accuracy, authority, aesthetics, professionalism, popularity, currency, impartiality, and quality of the information, can help people better assess web credibility.

As AGM emerged as a prominent technological advancement, there remains a noticeable gap in research concerning the assessment of credibility for these artefacts. Notably, a recent study examined people's strategies for identifying deepfake videos that use surface video and audio cues, comparing messages conveyed by the person(s) in the video with personal knowledge and searching for external sources to figure out the authenticity of the video [54]. Yet, there is a need for further exploration into how people, especially young individuals, assess the credibility of AGM such as deepfakes.

2.2.3. Authenticity verification as an approach. Rather than the psychological approaches (HSM and ELM) discussed in previous sections, which primarily examine individual cognitive processes in credibility evaluation skills, this research adopts a socio-constructionist/socio-cultural approach. Here, the focus shifts towards understanding credibility evaluation as a socio-cultural practice, emphasising the significance of authenticity verification in contemporary information environments.

According to Hussin et al. [59], credibility is commonly defined as the authenticity of information. In this sense, while facing undistinguishable information, authenticity verification can be an approach to testing information credibility. As Peterson [60] argues, 'authenticity is a claim that is made by or for someone, thing or performance and either accepted or rejected by relevant others'. He lists several examples to further elaborate on this concept, including the authenticity of products claimed in contemporary mundane product marketing campaigns and the real identity presented to others [60].

Many previous studies on the authenticity verification of information have been conducted in the field of media and communication, particularly focusing on digital information from social media platforms. For instance, Nee [61] investigated the authenticity verification of youths from the Middle East and the United States and found that their actions included 'googling' the headlines of stories and reading other people's comments.

As Peterson [60] points out, authenticity can be socially and technologically constructed, and the salience of authenticity changes all the time. More specifically, he articulated that authenticity can be illustrated by people with different backgrounds and that technology-based environments can impact authenticity to some extent [60]. Hence, in this study, it should be noted that the authenticity of AI-generated videos is both socially and technologically shaped and that the salience of authenticity might vary by different participants.

Moreover, there is a lack of research on the authenticity verification of information delivered by AGM. As we deal with diverse information in this post-truth era shaped by advanced generative technologies, it is urgent to explore how people verify the authenticity of new digital information as an approach to evaluating information credibility.

3. Methods

3.1. Research design

This study applies a qualitative research strategy to examine young people's practices while confronting AGM. As Denzin and Lincoln [62] argue, qualitative research encompasses a series of 'interpretive, material practices' to 'make the world visible'. Similarly, Bazeley [63] further elaborates that qualitative research aims to underline 'observing, describing, interpreting and analysing the way that people experience, act on, or think about themselves and the world around them'. For this study, qualitative analysis is employed as a way to interpret and unfold participants' practices and thoughts with AI-generated videos.

Oriented by these approaches, interviews were conducted to collect empirical data. During data collection, semi-structured interviews based on a prepared interview outline were designed. Notably, to observe and understand participants' ways of verifying the authenticity of videos, two AI-generated videos were provided for students to watch. The participants were allowed to make their judgements based on these video contents and share any thoughts on their minds with the researchers.

3.2. Participant recruitment, data collection and analysis

The participants' recruitment and data collection (which will be elaborated on in the next section) were conducted in collaboration with researchers from three departments at University A, Finland, in 2021. The data collection took place as part of the participants' work internships at the university. In Finland, a so-called work practice programme is involved in students' formal K-9 education. It entails students having a 1- to 2-week internship period in a workplace fulfilling certain requirements. The participants were recruited from two different classes in School B in Finland. Three senior researchers and a doctoral researcher from the research group visited the school to give a presentation on the internship opportunity at the university and introduce the associated research. Twenty-one ninth-grade students, aged 14 to 15 years, expressed their interest in taking part in the 1-week internship and participating in this study. The students who agreed to join the study and their legal guardians were asked to fill out consent forms. They were allowed to withdraw their consent at any time without impacting their participation in the internship.

The internship week offered students an opportunity to get to know the university as a working environment and to better comprehend the researchers' work. During the week, the participants engaged in various activities related to robotics, AI, and business. The activities consisted of learning about AI and robotics, thinking about AI-related business plans and participating in research-related activities. The internship occurred in a classroom-like environment at the local university.

Twenty participants ($N = 20$) were involved in the interviews, which followed a designated interview guide including six sections. After offering background for the interview and themes to be discussed, the main body of the interview involved the themes of 'AI in everyday apps', 'AI and personalised content', 'AI in finding media content', 'synthetic media / AI-generated content' and 'impacts of AI on young people'.

This research only focused on participants' responses under the theme of 'synthetic media / AI-generated content'. During this part of the interview, two AI-generated videos were provided for the participants to watch as an experiment, aside from interviews, driving young people to offer ideas about their ways of verifying the authenticity of the two videos and their thoughts about AI-generated videos. The two videos were prepared before the data collection. The sampling videos were selected from the README.md page of DeepFaceLab [64], a leading software for creating deepfakes, from the platform GitHub to ensure that the selected videos were generated by AI. There were six videos generated by DeepFaceLab listed under the title of 'What can I do using DeepFaceLab?', including De-Ageing With Deepfakes (Machine Learning), Примерим Навального в качестве президента, Рогозин ответил Илону Macky [DeepFake], Datazucc, Part 1 [deepfake], Deepfake Queen: 2020 Alternative Christmas Message and Dictators – Kim Jong-Un. Finally, two videos, Datazucc, Part 1 [deepfake] [65] and Dictators – Kim Jong-Un [66], were selected as the sampling videos. Other videos were expelled for reasons such as revealing information about 'deepfake' in the main body of the video and explicitly showing the scene of a 'changing face'.

The videos represented different scenarios, including political speech and science fiction (sci-fi) settings, and neither implied their property as a deepfake in their main body. As such, these two videos were regarded as appropriate material for the experiment. Once they finished watching each video, the participants were asked to consider whether the video they watched was authentic and how they identified it. After watching both videos, the interviewer explained that the videos were so-called deepfakes, explaining the meaning of the term. Subsequently, the participants were asked to reflect on how they felt about the videos, knowing that they were deepfakes and their thoughts about these two videos. All the interviews were conducted in English and recorded with the participants' permission.

All the empirical data from the interviews were first transcribed and then inductively coded. The transcription was segmented into every sentence as a unit to be analysed. Based on the analysis, seven categories were inductively extracted: 'image, sound, narrative, intuition, the techniques, the consequences and the associated people'. Furthermore, two central themes were identified from the categories: 'the strategies of authenticity verification' and 'the sense of mis- and disinformation'. Through this process, a coding frame was finalised, as shown in Table 1.

3.3. Ethical considerations

This research follows the ethical guidance released by the Finnish National Board on Research Integrity [67]. The empirical data collected with interviews can be considered personal data, since it probably makes the respondents identifiable. Therefore, the data have been protected in accordance with the General Data Protection Regulation [68]. The purposes for data usage were defined before the research began. Moreover, a detailed data management plan and privacy notice concerning the data collection were created.

During recruitment, the potential participants were told the purposes of this research and the data that would be collected. Moreover, before embarking upon the interviews, we offered consent forms that included the details of this

Table 1. The coding form of this study.

Central themes	Categories
The strategies of authenticity verification	Image
	Sound
	Narrative
	Intuition
The sense of mis- and disinformation	The techniques
	The consequences
	The associated people

research, promised the anonymisation of participants' information and clarified that data would be merely used for this research. Written consent was obtained from both the participants and their guardians.

4. Results

The first video that showed a political speech given by Kim Jong-Un was not identified as authentic by any of the participants. Instead, most of the participants considered it inauthentic or utilised other expressions to doubt the authenticity of the videos. A few informants answered that synthetic techniques might be applied to create the video. One person merely pointed out that the video was problematic instead of directly commenting on its authenticity. A quarter of the participants expressed uncertainty about the authenticity of the video.

The second video, which showed Mark Zuckerberg stating fake news issues in a spaceship-like setting, resulted in more varied takes on its (in)authenticity. Some participants thought that the science fiction setting of the video was inauthentic, while others pointed out that the characters in the video looked 'fake' or that the video might be synthesised by AI technologies. Three participants were unsure about the authenticity of the video. Only one person thought it might be authentic.

Based on the analysis, the first central theme, 'the strategies of authenticity verification', reveals the participants' practices to verify the authenticity of the videos from four perspectives: image, sound, narrative, and intuition. Rather than asking for the sources of the videos or looking for additional information, the judgements were primarily based on the video content – that is, message credibility [38]. Moreover, the three categories under the second central theme, 'the sense of mis- and disinformation', imply that the participants were able to sense mis- and disinformation in the context of AGM from three perspectives: the techniques, the consequences, and the associated people, as their reflections of the implications of AI-generated deepfake videos.

4.1. Young people's strategies of authenticity verification

Although the two AGM videos differed in terms of their scenarios, most participants focused on the visual and narrative aspects of their authenticity verification practices. Over one-fourth of the participants also took sound as a crucial perspective to verify authenticity. Merely, a small number of participants cannot elaborate on their practices to do so.

4.1.1. Image. For over half of the participants, visual information delivered by the images of both videos was a vital criterion to verify their authenticity. One perspective widely mentioned by the participants was the visual features of the two main characters presented in the videos. Their judgements were primarily oriented by the facial features of the characters. For example, some participants carefully observed the characters' faces, eyes, and mouths. After watching Video 1, a participant pointed out something strange with Kim Jong-Un's eyes: 'He wasn't blinking' (C8). Similarly, for Video 2, a participant described the appearance of the character in the video: 'He didn't look like a human. He looked more like a lizard person' (G19). In addition, young participants described their observations of how different parts of the character's body were combined. Examples such as B5's identification indicate this aspect:

> The face swapper and, most of the time, it stops at the neck. It's usually around here, the neck and all the way up. But then, if you look at around the neck ... and especially on the chest, it's still kind of fake, definitely (B5).

Beyond looking at the characters' body features, observing their body movements was another strategy for verifying authenticity. Some participants caught the subtly suspicious body movements of the two characters and thereby thought the two videos were inauthentic. For instance, E13's reply to Video 1 was, 'The movement looked odd' (E13). Likewise, G21 expressed the view, 'Because the way like he was moving was kind of weird' (G21). Also, for Video 2, B6 expressed his doubt: 'And also the movement of his mouth, it's different. [...] His mouth movement doesn't match with what he's saying' (B6).

The third perspective of the video image used by the participants for authenticity verification was the scene of the video. Three participants concentrated on the background of the scene. One of the representative answers was given by B5 for Video 1: 'And I'm guessing the background is green screen, because it looked kind of green screen' (B5). Compared with Video 1, the scene in Video 2 was more commonly noted. The same person, B5, noted, 'It's all CGI. It's mostly stuff for all green screens. It's like there's no way that's filming right in space' (B5). In addition, some judgements were based on the specific details of the scene. For example, E14 emphasised, 'Because we don't have those kinds of, like, spaceships or televisions yet ...' (E14).

Finally, a couple of participants mentioned other general video features. In this way, they did not point out any specific features in the video but relied on the general video image to verify its authenticity. For Video 1, one participant expressed, 'Doesn't really look real, like a deepfake or something. [...] It doesn't really look natural or kind of ... look right' (D11). Regarding Video 2, many participants shared their ideas about the whole video content. For example, F18 elaborated on this view: 'To me, it just seemed like a movie or something, not what's actually like happening now' (F18).

4.1.2. Sound. Around a quarter of the participants made their judgements based on the sounds of the two videos. More specifically, they primarily paid attention to the characters' voices. For Video 1, several participants pointed out the weirdness of Kim Jong-Un's voice. For example, B5 expressed to researchers, 'So I'm guessing that they use a face swapper and audio-like change thing' (B5). Another interviewee also noticed the character's voice and admitted that the judgements were made based on the sound (D12).

While verifying the authenticity of Video 1, one participant even took out his mobile phone to browse the videos he had watched in the past and elaborated on the assessment made from the perspective of sound: 'I tried to find a video he was speaking [...] I don't think that's a real voice. [...] I think they just made that voice sound more like ... more aggressive' (G19). Regarding Video 2, only two people assessed the authenticity of the video based on the sound, mainly focusing on the main character's voice. As one participant argued, 'The voice [...] changes, like I'm pretty sure that voice is the voice of the CEO of Facebook'. (B5) In line with B5's opinion, C7 stated that the main character's voice 'sounded a lot like the actual CEO [of Meta]' (C7).

4.1.3. Narrative. Notably, for the video presenting a non-fiction setting, Video 1, the participants were found to analyse the video's narrative. By comparing the narratives with their existing knowledge, they could determine whether the video content made sense. Two participants doubted whether Kim Jong-Un could speak English as fluently as the video shows. Two examples reveal this point: 'Kim Jong-Un, can he speak fluent English?' (A1) and 'I don't know how well he speaks English, but I don't think he knows how to speak English that well' (C7).

Some noticed the words spoken by Kim Jong-Un. In this video, Kim Jong-Un, a national leader of North Korea, consistently stated that democracy is vulnerable and easily manipulated by people, which is generally a scenario that is unlikely to occur. As A2 said, 'It's just that the context of him speaking about democracy and him posting that, it's just weird' (A2). In line with A2, F17 articulated:

I feel like Kim Jong-Un isn't the type of person to voice out his political opinion [...] I've never seen before that it was like, 'Oh my gosh, democracy is bad ...' (F17).

For Video 2, one person identified that the narrative was confusing and thereby verified the video as inauthentic: 'I don't know ... And the same dude was in bed or something weird. I don't know what's going on' (A3).

4.1.4. Intuition. Finally, only four students expressed that they verified the authenticity of the videos but were not able to explicitly explain their strategies for doing so or the basis for supporting their judgements. Here, we have labelled this strategy as 'intuition'. One of the participants offered the identification of authenticity without a specific reason for Video 1:

Researcher: Is there something that you could say that this is how I know that it's not real?
A3: I don't know.

Researcher: Just a feeling?

A3: Yeah.

Another participant, F18, was trying to find evidence to back up the authenticity verification: 'I don't know, it was pretty weird. [...] I don't know ...' (F18). For Video 2, two participants' reactions were: 'I don't know how to explain it. I just like ... question it' (C7) and 'I don't know. It just like ... didn't seem true' (E15).

4.2. Young people's sense of mis- and disinformation in the context of AGM

After watching the two videos and discussing them, we explained to the participants that they were so-called deepfake videos and what that meant. Then, we asked the participants about their feelings regarding the videos, knowing that they were deepfakes, and their thoughts about the content and quality of the two videos. This discussion resulted in reflections about the implications of deepfake videos beyond what they had just observed, centring on their sense of mis- and disinformation in the context of AGM from three perspectives: (1) techniques; (2) consequences, and (3) associated people.

4.2.1. Sense of techniques. During the discussion, participants generally provided their thoughts on these two videos instead of separating them. Their reflections, based on the content and quality of the videos, revealed their sense of the techniques underlying AGM. Many of the students felt that the quality of these videos was good or that they looked very real, which revealed that the techniques hidden behind the two videos were mature. For example, A1 repetitively utilised descriptions such as 'it's hard to like tell if it's fake', 'the quality is really well [done]' and 'the quality is so good' (A1) to emphasise the advanced techniques. Others also shared similar opinions, as B5 stated:

> [...] Because it looks really real, like I act, like if you were probably giving that to someone who was young and had no idea about AI, then they would probably believe that that's actually that person. (B5)

A few participants offered a more detailed assessment of AGM's techniques based on a comparison between the two videos. For instance, B6 thought that Video 1 was really hard to tell that 'it was fake', but Video 2's quality was not that good. D10 held similar views to B6 and articulated them with richer and more critical judgements. According to D10's response, Video 1's image 'looks authentic'; however, the sound is 'not like very well executed' (D10); in terms of Video 2, D10 considered the content to be 'completely [...] not professional', but the main characters' 'shadows', 'movements' and 'lip syncing' were 'quite better' (D10).

Moreover, several participants showed the same views as the aforementioned ones. Participant D11 actually realised that the videos were deepfakes during the watching process, even though admitting the quality is good but can be better. Both E14 and E15's responses indicated the immaturity of the techniques used in these videos. E14 signified, 'I don't know. It doesn't surprise me, because they didn't look real at all' (E14). E15 considered Video 1 to be 'obviously fake' and 'did not think there is any hard' to verify its authenticity (E15).

Beyond looking at the current techniques applied by AGM, seven participants shared their thoughts about future techniques by predicting progress in the quality of deepfake videos. For instance, as E13' assessed, 'It'll probably get even better quality, and then it's going to be way harder to distinguish what's real and what isn't' (E13). The opposite views existed as well. For instance, F17 thought the content creators 'could do a lot better with the editing' (F17).

4.2.2. Sense of consequences. Participants considered that deepfakes delivered mis- and disinformation. Based on this perception, over three-fifths of the participants mentioned the consequences that might be brought about by deepfake, with most of them being negative. According to their reflections, the negative consequences primarily included issues related to security, democracy and bias.

Regarding issues related to security, a small number of the participants expressed concerns about identity theft and privacy. These interviewees particularly centred on the methods of face and voice 'stealing'. One of the representative responses came from B5:

Well, I mean especially in the future, it's gonna get more real, right, which is really bad because identity theft is a really big thing [that] could happen in the future, with identity theft of ... because you know the face is probably the one thing that you're gonna see, is locked onto your identity card and everything that all has your face (B5).

According to their elaboration, someone's identity could be stolen by using deepfake techniques, which may cause security problems. F17 also thought that 'someone else is using their face to talk like they're using a voice changer to make their voice sound like this to pretend that they are actually them' (F17). In line with B5 and F17's responses, C8

noted that by using filters or other AGM, a person can change into a different person on the Internet, even taking some-one else's identity (C8).

Second, some of the participants discussed issues related to democracy as a consequence of AGM. Some combined the consequences of fake news, as F16 emphasised: 'Fake news can like very well spread from that'. Another person pointed out that AGM could be used in disinformation campaigns, which might impact democracy (D10). Apart from these two responses, F17 offered a perspective on 'conspiracy theories'. More specifically, this participant noticed a stereotypical pattern in the deepfake videos they had encountered: 'If you search conspiracy theories, a lot of videos are like those' (F17), which might cause threats to democracy.

Third, one participant expressed that this kind of video could result in bias by declaring that the deepfake videos might make 'somebody's own personal life ... get biased' (D12). From their point of view, these videos might build stereotypes about someone's life. Hence, D12 determined that creating these videos is 'not the right thing to do'.

Beyond talking about the negative consequences of deepfake videos, a few participants showed their negative emotions towards the consequences brought by AGM. They used strong words, such as 'alarming', 'scary' and 'suspicious' to describe deepfake videos. For example, B5 and C8, respectively, characterised these videos as 'kind of alarming' (B5) and 'kind of scary' (C8) after watching the samples. What's more, C9's reply was to point to general distrust that deep-fake videos may be raising: 'I feel a bit suspicious about some videos, it's like ... I can't trust every video I see' (C9).

Some of the participants, even though they did not scrutinise the concrete consequences of the mis- and disinformation of AGM, affirmed that this mis- and disinformation could harm society. For instance, E13 took Video 1 as an example: 'The first one could probably be used to ... like spread a lot of misinformation. So, it's kind of harmful that, you know, deepfakes are, like, spreading around' (E13).

Notably, only two of the participants talked about the more favourable consequences of creating an AGM. E13 suggested, 'When [deepfakes are] not taken too seriously, like when it's just a joke, then it can be alright' (E13), while E14 pointed out entertainment as the usage of deepfake videos.

4.2.3. Sense of associated people.
In addition to talking about the techniques behind AGM and the consequences of AGM, participants also articulated their thoughts about people who are associated with AGM, primarily involving users of AGM, the receivers of AGM and other people who are impacted by AGM. Similar to the results in the previous section, they mostly offered negative assessments.

While talking about users who can make AGM, participants' responses were primarily directed towards malicious and abusive use of AGM. Over one-fourth of the participants articulated this from the perspective of the AGM creators. For instance, A2 described the possibility: 'Someone could probably easily use it to start a conflict between two people by using the filter and then saying something and then posting it' (A2). Other participants offered their views from the side of AGM manipulators, who may not create AGM but utilise them. For example, E15 claimed worries about people who have power: 'So that's kind of a problem that people can have the power to kind of do that stuff' (E15).

In addition, a few interviewees discussed the challenges confronted by the receivers of AGM, as well as others who may be impacted by AGM. Here, they mostly regarded these relevant people as laypeople who were not professionals in the field of AI or who had powerful MIL. In this context, participants such as B5, people who are young or have no knowledge of AI, might be fooled by AGM. Likewise, D10 declared that people 'who are not aware of this technology' might take the content of AGM as truth (D10). According to F16, some people might believe in AGM. In line with B5, D10 and F16, F18 took the angle of the first video to share experiences as an AGM receiver:

> I mean, the first one. I could tell something was off, but I mean, I didn't know that someone had actually pasted it on someone else's face. I didn't realise that. [...] It's just not good if it is that easy to miss. (F18)

In terms of others who are impacted by AGM but might not be receivers, B5 stated a potential situation: 'So if you can change your face into someone else's like an identity theft, it is gonna be way easier, and it's going to be a way bigger problem' (B5). From this possibility raised by the participants, it can be seen that people who are not even directly associated with AGM may be harmed by them.

Overall, this study showed Finnish young people's strategies for verifying the authenticity of AGM content and their sense of mis- and disinformation in the context of AGM. When assessing the authenticity of the two provided videos, young Finns primarily based their judgements on four key elements: image, sound, narrative and intuition. In addition, their reflections on the video content highlighted a sense of mis- and disinformation from three perspectives: the techniques related to AGM, the consequences of AGM and people associated with AGM.

5. Discussion

5.1. Young Finns' MIL regarding deepfakes

The results of this study pointed to the participants' strategies of authenticity verification regarding AGM from four perspectives, namely image, sound, narrative, and intuition. While comparing with existing research on young people's strategies for credibility assessment, some overlaps can be identified. For example, participants in this research paid attention to the texts presented in two videos, which can be regarded as similar strategies to eighth-grade students' credibility assessment of unfamiliar websites in the study by Kohnen et al. [56].

Surprisingly, an examination of the strategies used by young Finns in our study reveals that they possess capabilities similar to those of adults, as demonstrated in Goh's [54] research. This includes utilising surface video and audio cues and comparing the messages conveyed by individuals in the video with their personal knowledge.

In addition to young Finns' credibility assessment of deepfakes, it can be found that young people have some basic knowledge related to AGM or AI, which is shown in their senses of mis- and disinformation. This indicates that their MIL is not merely about credibility assessment but also involves a fundamental comprehension of the technologies that produce such content. As AI technologies continue to evolve and become more pervasive, their role in shaping MIL cannot be overlooked. The integration of technological literacy, particularly regarding AI, is essential in equipping young people with the skills to critically assess and navigate complex digital environments.

5.2. Message credibility matters

Whereas 'relying on affiliation with the source' was one of the key methods of authenticity verification in Waruwu et al.'s [69] research, sourcing was not mentioned by any of the participants in our research. During the data collection, we did not restrict the participants' ways of verifying the authenticity of the videos and allowed them to share their thoughts or ask any questions. However, none of them asked questions about the sources of the videos or talked about the perspective of sourcing. Instead, all participants based their authenticity verification directly on the messages themselves, encompassing the characteristics of the message, such as 'content, structure, language and presentation' [38]. Hence, this study shows how message credibility matters for assessing information credibility.

Regarding authenticity verification based on the messages of two videos, the findings show that while making the authenticity verification, the young participants performed both explicit and implicit comparisons between the message and their existing knowledge. Explicit comparisons were presented in expressions in which the details of the process of authenticity verification were explained. It is impressive that many young participants had basic knowledge of Kim Jong-Un, Mark Zuckerberg, and Star Wars as key elements of the videos offered by us and knowledge of AI and film production (e.g. green screen and computer-generated imagery techniques). To some extent, this existing knowledge influenced their judgements of authenticity. Implicit comparisons were exemplified by expressions directly pointing out the so-called 'weird parts' of the videos, such as abnormal movement, primarily revealed in the strategy based on their intuition. It is interesting that, despite being oriented by intuition, these participants could feel the inauthenticity of the videos.

Moreover, to assess message credibility, some participants attempted to build triangulation to solidify their authenticity verification based on the video [70]. For example, one participant pointed out the strangeness of both the movements of the character's face and neck and the scene of the video to verify the inauthenticity of the video content.

5.3. Constructed authenticity of AGM

The findings of this study indicate that, from the perspective of its young participants, authenticity mostly means the real thing, the fact, the truth and the non-fictionality of the scenario. However, the notions of authenticity verification varied based on the genres and contexts of the videos. The two deepfake videos used in this study prompted a discussion of different notions of authenticity. First, in the political speech-like video, authenticity was primarily understood to refer to the speech as true and as something that really happened. Second, in the sci-fi-like video, viewers' attention to authenticity seemed to be switched to deliberations about whether the narrative was fictional or nonfictional, whether the scene was real or created using technology, and whether the character was a real human.

According to Peterson [60], the salience of authenticity varies among individuals and within technology-influenced environments, remaining dynamic over time. In this study, youths articulated their assessments of the believability of video content based on different forms of evidence. This variation in judgement reflects the notion that authenticity is not a fixed concept but is constructed in multiple ways. This finding aligns with Peterson's [60] argument that authenticity can be understood through diverse contexts. Although the young participants' ideas of authenticity varied, their notion of

inauthenticity was rather uniform – almost every participant considered inauthenticity negatively, demonstrating a high level of consistency. Furthermore, deepfake videos were mostly considered negative artefacts by the young participants. However, this negative take does not consider the potential positive benefits AGM can bring to society. For example, AGM has been used to create interactive components in museums, and as such can deliver diverse aesthetic values in the film industry to recreate classic scenes without refilming them, in the game industry to enable multiplayer games and virtual chat worlds with increased telepresence, and even to help digitally bring a deceased person 'back to life' to allow people to farewell or memorialise their loved ones [2,4,71].

5.4. Evaluation of method and future research

This research applied a qualitative approach to investigate young people's authenticity verification processes and the implications of AGM, which allowed us to gain rich and detailed insights from conversations with the participants. Moreover, offering videos as an experiment was helpful in elucidating the participants' processes of authenticity verification.

Naturally, the methods employed have limitations. First, as our video selection was within limited options based on the GitHub page of DeepFaceLab, the final sample of videos did not cover many different genres or topics of deepfake videos. Thus, for future research, the samples of videos could be expanded, with an attempt to involve more variety in the genres and topics of AGM. In this way, we may be able to gain more nuanced insight into the people's reflections on various video materials.

Second, the scenario we set for interviewees to verify authenticity was not representative of the settings of their everyday lives. In a natural setting, they might not verify the authenticity of every piece of information they come across, or they might have sought background information. Hence, in the scenario built by us, their reactions may be different from those in real life. Regarding future research, an ethnographic approach can be valuable in capturing how people deal with AGM in their everyday lives.

6. Conclusion

This study contributes to the exploration of AGM from the perspective of young people's MIL practices. By applying a qualitative approach, this study investigates young people's practices to evaluate the content of AI-generated videos. The results present young Finns' strategies of authenticity verification regarding AGM from four perspectives, based on image, sound, narrative, and intuition, and how they sense mis- and disinformation from three perspectives, including a sense of techniques hidden behind AGM, consequences of AGM and people associated with AGM. The research can enhance awareness that encourages potential remedies in terms of MIL education or policy initiatives that take the problematic consequences of AGM into account.

ORCID iDs

Yucong Lao https://orcid.org/0000-0003-1979-8748
Noora Hirvonen https://orcid.org/0000-0001-7577-3057

Notes

1. DALL-E 3 is an advanced version of the DALL-E, a text-to-image system, capable of generating images based on user-input prompts.
2. ChatGPT is a chatbot, based on large language models, that enables users to print in prompts to get humanlike responses generated by the system.
3. OpenAI is an artificial intelligence research organisation known for developing advanced AI technologies, including signature products like ChatGPT and DALL-E.

4. Sora is an AI model capable of generating realistic and imaginative scenes based on users' text instructions.
5. American actor and director Jordan Peele used deepfake technology to map his own facial movements onto former U.S. President Barack Obama's face, creating a video.

References

[1] Partadiredja RA, Serrano CE and Ljubenkov D. AI or human: the socio-ethical implications of AI-generated media content. In: *2020 13th CMI conference on cybersecurity and privacy (CMI)-digital transformation-potentials and challenges (51275)*. Copenhagen: Denmark, 26–27 November 2020, pp. 1–6. New York: IEEE.

[2] Mihailova M. To dally with Dalí: deepfake (Inter) faces in the art museum. *Convergence* 2021; 27(4): 882–898.

[3] Sample I. What are deepfakes – and how can you spot them? *Guardian*. https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them (2020, accessed 26 March 2023)

[4] Westerlund M. The emergence of deepfake technology: a review. *Technol Innov Manag Rev* 2019; 9(11): 39–52.

[5] Sayler KM and Harris LA. Deep fakes and national security. *Report, congressional research SVC Washington, United States*, August 2020.

[6] Vaccari C and Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty and trust in news. *Soc Media Soc* 2020; 6(1): 1–13.

[7] Goodfellow IJ. General adversarial nets. *arXiv*. 2014; 3(11): 139–144.

[8] De Vries K. You never fake alone. Creative AI in action. *Inf Commun Soc* 2020; 23(14): 2110–2127.

[9] Kolo C, Mütterlein J and Schmid SA. Believing journalists, AI, or fake news: the role of trust in media. In: *Hawaii international conference on system sciences (HICSS)*, Maui, USA, 4–7 January 2022, pp. 1–10. ScholarSpace.

[10] Hofeditz L, Mirbabaie M, Holstein J, et al. Do you trust an AI-journalist? A credibility analysis of news content with AI-authorship. In: *European conference on information systems*, Marrakech, Morocco, 14–16 June 2021, pp. 1–15, paper no. 1272.

[11] Whittaker L, Kietzmann TC, Kietzmann J, et al. 'All around me are synthetic faces': the mad world of AI-generated media. *IT Prof* 2020; 22(5): 90–99.

[12] Ajder H, Patrini G, Cavalli F, et al. The state of deepfakes: landscape, threats, and impact. Report, DEEPTRACE, Milan, Italy, September 2019.

[13] Langa J. Deepfakes, real consequences: crafting legislation to combat threats posed by deepfakes. *BU L Rev* 2021; 101: 761–801.

[14] Ali S, DiPaola D, Lee I, et al. Children as creators, thinkers and citizens in an AI-driven future. *Comput Educ Artif Intell* 2021; 2: 100040.

[15] Auxier B and Anderson M. Social media use in 2021. Report, pew research center, US, April 2021.

[16] Satariano A. Young people more likely to believe virus misinformation, study says. *The New York Times*. https://www.nytimes.com/2020/09/23/technology/young-people-more-likely-to-believe-virus-misinformation-study-says.html (2020, accessed 23 September 2020)

[17] Guess A, Nagler J and Tucker J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Sci Adv* 2019; 5(1): eaau4586.

[18] UNESCO. Global standards for media and information literacy curricula development guidelines. Report, UNESCO, Republic of Serbia, September 2019.

[19] Charron N and Rothstein B. Social trust, quality of government and ethnic diversity: an empirical analysis of 206 regions of Europe. Report, QoG Working Paper Series, Gothenburg, December 2014.

[20] OECD Building Trust in Public Institutions. *Drivers of trust in public institution in Finland*. Paris: OECD Publishing, 2021.

[21] Karjalainen A. Digitalisation is a key enabler for public sector renewal. *State Treasury Republic of Finland*. https://www.treasuryfinland.fi/annualreview2020/digitalisation-is-a-key-enabler-for-public-sector-renewal/ (2020, accessed 26 April 2023)

[22] Lessenski M. How it started, how it is going: media literacy index 2022. Report, Media Literacy Index project of the European Policies Initiative (EuPI) of the Open Society Institute–Sofia Foundation (OSI-Sofia), Sofia, Bulgaria, October 2022.

[23] Mackintosh E. Finland is winning the war on fake news. *What it's learned may be crucial to Western democracy*. CNN. https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/ (2019, accessed 26 March 2023)

[24] Frau-Meigs D. Information disorders: risks and opportunities for digital media and information literacy? *Medijske Studije* 2019; 10(19): 10–28.

[25] Singh N and Banga G. Media and information literacy for developing resistance to 'infodemic': lessons to be learnt from the binge of misinformation during COVID-19 pandemic. *Media Cult Soc* 2022; 44(1): 161–171.

[26] UNESCO. MIL curriculum and competency framework for teachers. UNESCO. http://en.unesco/media-and-information-literacy-curriculum-for-teachers (2012)

[27] Frau-Meigs D. Media and information literacy. Report, paper commissioned for the 2023 Global education monitoring report, technology in education. Document code: ED/GEMR/MRT/2023/P1/06, UNESCO, France, 2023.

[28] Grizzle A, Moore P, Dezuanni M, et al. Media and information literacy: policy and strategy guidelines. Report, UNESCO, Paris, 2013.

[29] Haider J and Sundin O. *Paradoxes of media and information literacy: the crisis of information*. London: Routledge, 2022.

[30] Brisola AC and Doyle A. Critical information literacy as a path to resist 'fake news': Understanding disinformation as the root problem. *Open Inf Sci* 2019; 3(1): 274–286.

[31] Abu-Fadil M.Combatting disinformation and misinformation through Media and Information Literacy (MIL). In: Ireton C and Posetti J eds. *Journalism, 'fake news' and disinformation: a handbook for journalism education and training*. Paris: UNESCO, 2018, pp. 70–80.

[32] Tandoc Jr EC, Lim D and Ling R. Diffusion of disinformation: how social media users respond to fake news and why. *Journalism* 2020; 21(3): 381–398.

[33] Reem M. The impact of media and information literacy on students' acquisition of the skills needed to detect fake news. *J Media Lit Educ* 2022; 14(2): 58–71.

[34] Lee Y, Huang KT, Blom R, et al. To believe or not to believe: framing analysis of content and audience response of top 10 deep-fake videos on youtube. *Cyberpsychol Behav Soc Netw* 2021; 24(3): 153–158.

[35] Wilson C, Grizzle A, Tuazon R, et al. Media and information literacy curriculum for teachers. Report, UNESCO, Paris, 2011.

[36] Rieh SY and Danielson DR. Credibility: a multidisciplinary framework. *Annu Rev Inform Sci Technol* 2007; 41(1): 307–364.

[37] Tseng S and Fogg BJ. Credibility and computing technology. *Commun ACM* 1999; 42(5): 39–44.

[38] Rieh SY.Credibility and cognitive authority of information. *Encyclopedia of library and information sciences*. New York: Taylor and Francis Group, 2010, pp. 1337–1344.

[39] Appelman A and Sundar SS. Measuring message credibility: construction and validation of an exclusive scale. *J Mass Commun Q* 2016; 93(1): 59–79.

[40] Febriane R, Wibowo W and Agrippina YR. The influences of perceived credibility and consumer attitude towards purchase intention of some by Mi's YouTube user generated content. In: *4th Asia Pacific management research conference (APMRC 2022)*, Jakarta, Indonesia, 18–20 May 2022, pp. 324–337. Dordrecht, The Netherlands: Atlantis Press.

[41] Shariff SM. A review on credibility perception of online information. In: *2020 14th international conference on ubiquitous information management and communication (imcom)*, Taichung, Taiwan, 3–5 January 2020, pp.1–7. New York: IEEE.

[42] Metzger MJ, Flanagin AJ and Zwarun L. College student web use, perceptions of information credibility, and verification behavior. *Comput Educ* 2003; 41(3): 271–290.

[43] Courtright C. Context in information behavior research. *Annu Rev Inform Sci Technol*. 2007; 41(1): 273–306.

[44] Rieh SY and Hilligoss B.College students' credibility judgments in the information-seeking process. *Digital media, youth, and credibility*. Cambridge, MA: The MIT Press, 2008, pp. 49–72.

[45] Kiousis S. Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Commun Soc* 2001; 4(4): 381–403.

[46] Li R and Suh A. Factors influencing information credibility on social media platforms: evidence from Facebook pages. *Procedia Comput Sci* 2015; 72: 314–328.

[47] Metzger MJ, Flanagin AJ, Eyal K, et al. Credibility for the 21st century: integrating perspectives on source, message, and media credibility in the contemporary media environment. *Ann Int Commun Assoc* 2003; 27(1): 293–335.

[48] Gass RH and Seiter JS. *Persuasion, social influence, and compliance gaining*. Boston: Allyn & Bacon, 1999.

[49] Perloff RM. *The dynamics of persuasion*. Hillsdale: Lawrence Erlbaum Associates, Inc, 1993.

[50] Aharon AA, Ruban A and Dubovi I. Knowledge and information credibility evaluation strategies regarding COVID-19: a cross-sectional study. *Nurs Outlook* 2021; 69(1): 22–31.

[51] Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J Pers Soc Psychol* 1980; 39(5): 752–766.

[52] Chen S and Chaiken S.The heuristic-systematic model in its broader context. In: Chaiken S and Trope Y eds. *Dual-process theories in social psychology*. New York: The Guildford Press, 1999, pp. 73–96.

[53] Petty RE and Cacioppo JT. *The elaboration likelihood model of persuasion*. New York: Springer, 1986.

[54] Goh DHL. 'He looks very real': media, knowledge, and search-based strategies for deepfake identification. *J Assoc Inf Sci Technol* 2024; 75(6): 643–654.

[55] Subramaniam M, Taylor NG, St Jean B, et al. As simple as that? Tween credibility assessment in a complex online world. *J Doc* 2015; 71(3): 550–571.

[56] Kohnen AM, Mertens GE, Dawson K, et al. A study of middle school students' online credibility assessments: challenges and possibilities. *Res Teach Engl* 2021; 56(1): 33–59.

[57] Lederman R, Fan H, Smith S, et al. Who can you trust? Credibility assessment in online health forums. *Health Policy Technol* 2014; 3(1): 13–25.

[58] Shah AA, Ravana SD, Hamid S, et al. Web credibility assessment: affecting factors and assessment techniques. *Inform Res* 2015; 20(1): 1–28.

[59] Hussin N, Hashim H, Mokhtar WNHW, et al. Theorizing online information credibility among teenagers. *Int J Acad Res* 2020; 10(11): 516–526.

[60] Peterson RA. In search of authenticity. *J Manag Stud* 2005; 42(5): 1083–1098.

[61] Nee RC. Youthquakes in a post-truth era: exploring social media news use and information verification actions among global teens and young adults. *J Mass Commun Educ* 2019; 74(2): 171–184.

[62] Denzin NK and Lincoln YS.The discipline and practice of qualitative research. In: Denzin NK and Lincoln YS eds. *Handbook of qualitative research*. US: Sage, 2005, pp. 1–43.

[63] Bazeley P. *Qualitative data analysis: practical strategies*. US: Sage, 2013.

[64] Deepfacelab. GitHub. https://github.com/iperov/DeepFaceLab (accessed 14 December 2021)

[65] Datazucc. Part 1 [deepfake]. YouTube. https://www.youtube.com/watch?v=R9f7WD0gKPo (2020, accessed 14 December 2021)

[66] Dictators. Kim Jong-un. YouTube. https://www.youtube.com/watch?v=ERQlaJ_czHU (2021, accessed 14 December 2021)

[67] TENK. The ethical principles of research with human participants and ethical review in the human sciences in Finland. TENK. https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf (2019, accessed 26 January 2025)

[68] European Parliament & Council of the European Union. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2016/679/oj (2016, accessed 23 January 2025)

[69] Waruwu BK, Tandoc Jr EC, Duffy A, et al. Telling lies together? Sharing news as a form of social authentication. *New Media Soc* 2021; 23(9): 2516–2533.

[70] Greyson D. Evolution of information practices over time. *Proc Assoc Inf Sci Technol* 2016; 53(1): 1–8.

[71] Larsson S. Necrorobotics: the ethics of personalised resurrection. In: *Companion of the 2023 ACM/IEEE international conference on human-robot interaction (HRI '23 Companion)*. Stockholm, Sweden, 13–16 March 2023, pp. 1–5.