



Deepfaking the past: Memory and perceived truth of resurrected historical figures

María T. Soto-Sanfiel ^{*} , Gina Junhan Fu

Department of Communications and New Media, Centre for Trusted Internet and Community, National University of Singapore, Singapore

ARTICLE INFO

Keywords:

Deepfakes
Historical narratives
Perceived truth
Recall
Modality
Need for cognition

ABSTRACT

Deepfakes have intensified fears about the power of synthetic media to reshape historical memory and blur the boundaries between truth and fabrication. This study compares how audiences process historical narratives presented by deepfaked figures and in the form of written text, focusing on differences in recall and perceived truth over time. In an online experiment, 332 U.S. participants read or watched first-person accounts by Albert Einstein or Marie Curie across two sessions one week apart. Written narratives produced higher recall and stronger perceived truth than deepfake videos, both immediately and one week later. Also, Need for Cognition predicted greater perceived truth across conditions, suggesting that analytic engagement rather than technological realism drives belief in mediated representations. Although perceived truth declined over time, the drop was steeper for deepfake messages. Age also emerged as a significant predictor of some measures of perceived truth and recall. Pairwise comparisons between age groups showed progressive increases with participants aged 18–29 displaying the weakest recall and truth ratings, whereas those aged 30 and above maintained stronger and more stable judgments over time, independent of analytic motivation. These findings indicate that analytically engaged audiences are resilient to synthetic realism, limiting the potential for deepfakes to distort historical understanding.

1. Introduction

Deepfakes are AI-generated media that are typically created with generative adversarial networks (GANs) and can convincingly reproduce the gestures, speech, and facial expressions of real people (Altuncu et al., 2024; Chesney & Citron, 2019). Once associated mainly with deception, fraud, and political disinformation (e.g., Vaccari & Chadwick, 2020), their uses have diversified. Increasingly, “benign” deepfakes appear in educational, commemorative, artistic, and prosocial contexts, leveraging the familiarity of well-known faces to evoke reflection, empathy, and social connection but without the intention to deceive (Lundberg & Mozellius, 2025; Murillo-Ligorred et al., 2023; Soto-Sanfiel et al., 2025). As realistic synthetic renderings of public figures spread, traditional cues of authenticity (e.g., visual fidelity, vocal match, or production context) are no longer reliable. Judgments of truth therefore depend less on sensory evidence and more on cognitive processes such as attention, memory, and belief, often driven by fluency and familiarity rather than visual realism (Brashier & Marsh, 2020; Fazio et al., 2015; Pennycook et al., 2018).

Among the most ethically charged forms of synthetic media are deepfakes of deceased individuals, a phenomenon known as digital resurrection (P’Ng, 2024). These recreations blend documentary reality with fiction, enabling posthumous communication through the likeness and voice of historical figures. Clearly benign examples of these deepfakes include a museum installation in which visitors are able to interact with Salvador Dalí (Mihailova, 2021) and the short movie *Event of Moon Disaster*, in which a synthetic Richard Nixon delivers the contingency speech prepared for the Apollo 11 mission (Lees, 2024). However, despite their benign nature, such portrayals raise not only emotional and ethical questions around authenticity and consent but also epistemic concerns about how deepfakes blur the line between factual record and creative speculation, reshaping historical truth and collective memory (American Historical Association, 2021). Audience research shows that responses to benign deepfakes of deceased figures vary across cultural contexts and depend on perceptions of realism, legacy, and moral acceptability (Soto-Sanfiel & Wu, 2026). However, there is still insufficient empirical evidence regarding how truthful these posthumous representations are perceived to be or how such judgments evolve over

^{*} Corresponding author.

E-mail addresses: cnmmmtss@nus.edu.sg (M.T. Soto-Sanfiel), gina.junhan.fu@u.nus.edu (G.J. Fu).

<https://doi.org/10.1016/j.chb.2026.109008>

Received 3 November 2025; Received in revised form 20 March 2026; Accepted 29 March 2026

Available online 1 April 2026

0747-5632/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

time.

The growing realism and accessibility of deepfakes have intensified anxieties about the erosion of truth and manipulation of history, undermining trust in mediated communication. Beyond political disinformation, deepfakes pose ontological questions about what counts as “real” when the past can be algorithmically recreated. However, while public debate often centers on deception and vulnerability, far less is known about how audiences actually perceive and remember synthetic narratives in comparison with traditional formats such as text. Understanding these cognitive processes is essential to determine whether deepfakes truly possess the persuasive and mnemonic power often attributed to them—or whether critically aware and media-literate audiences can resist their epistemic allure.

The present study investigates how perceived truth and memory change over time when deceased historical figures are portrayed through deepfakes. Using a longitudinal experimental design, we compared deepfake and written versions of identical first-person narratives delivered by Albert Einstein and Marie Curie to test whether perceptual realism enhances or alters belief and recall when participants are aware that the material is artificial. The study also examines whether individual differences in cognitive motivation moderate these effects. This research contributes empirical evidence to ongoing debates on the persuasive power of deepfakes, helping to explain whether the risks attributed to them stem from their deceptive potential or from the communicative affordances of the audiovisual form itself.

2. Literature review

2.1. From Defaults to Deception: Cognitive Foundations of Perceived Truth and Memory

Human beings are not objective evaluators of information. According to Truth-Default Theory (TDT; Levine, 2014, 2021), people usually operate under a presumption of honesty: we tend to believe messages are true unless specific cues trigger suspicion. This truth-default is socially adaptive because most communication is genuine, but it can also lead to vulnerability when deceptive messages offer no clear diagnostic signs. When viewers of deepfakes encounter realistic but non-diagnostic (i.e., non-revealing) stimuli, they may accept them at face value before skepticism arises, allowing fabricated material to be encoded and remembered as genuine (Levine, 2014, 2021).

Empirical research confirms that humans are remarkably poor at detecting synthetic media. People often struggle to distinguish AI-generated from authentic audiovisual material and have even been found to rate artificial versions as more trustworthy or convincing (Groh et al., 2024; Nightingale & Farid, 2022). Awareness labels such as “AI-generated content” tend to increase skepticism, but only modestly, with little effect on perceived accuracy or sharing intentions (Gamage et al., 2025; Li & Yang, 2024). Initial belief in deepfakes thus stems less from naïveté than from ordinary cognitive defaults interacting with perceptual realism.

Over time, early impressions may consolidate into more persistent judgments of truth and memory. Research on the illusory truth effect shows that repeated or fluid information is more likely to be remembered and believed, even when it is known to be false (Fazio et al., 2015; Hasher et al., 1977). The perceptual fluency and emotional vividness of audiovisual deepfakes may therefore amplify their cognitive impact compared with written narratives, particularly during the early stages of encoding. However, longitudinal evidence on the durability of such effects remains scarce, as most studies rely on short-term designs without delayed measurement (Murphy et al., 2023). Clarification of how perceived truth evolves after initial exposure is thus essential for determining whether synthetic realism produces transient or enduring distortions in belief and recall.

Cognitive research on false memory formation provides longitudinal evidence that the illusory truth effect extends to the fragility of memory.

Repeated exposure enhances processing fluency—the ease with which information is processed—which people often interpret as evidence of truth (Dechêne et al., 2010; Hasher et al., 1977; Fazio et al., 2015). Over time, detailed episodic traces fade while gist-based representations dominate (Reyna & Brainerd, 1995), weakening source monitoring and encouraging misremembering. Loftus's (2005) classic misinformation studies show that vivid but false material can distort recall, and more recent evidence confirms that audiovisual realism amplifies memory malleability (Deck & Paterson, 2020). Consequently, repeated or emotionally engaging deepfakes may consolidate not only belief but also false memories, leading their influence to persist beyond initial exposure.

Meta-analyses of deception detection show that average accuracy is little more than chance ($\approx 54\%$) and that people are typically better at recognizing truths than lies—a pattern known as the veracity effect (Bond & DePaulo, 2006). Likewise, Luo et al. (2022) found that individuals' ability to distinguish real from fake news headlines on social media also approximates chance, driven by pervasive truth bias and contextual cues such as endorsements (e.g., likes, shares, positive comments). However, these studies are largely cross-sectional and text-based, with little exploration of how audiovisual realism and temporal distance interact to influence truth judgments and memory consolidation.

Recent experiments indicate that deepfakes can distort or even generate memories of non-existent events (Murphy & Flynn, 2022). Most of this work, however, measures effects immediately after exposure, offering little insight into how beliefs develop as initial impressions fade. Reviews consistently note the scarcity of longitudinal evidence on the persistence or transformation of these effects (Murphy et al., 2025). But temporal distance is crucial: as time passes, individuals rely less on surface cues such as realism or production quality and more on source familiarity and interpretive context. Early impressions of authenticity may therefore be revised as audiences reflect upon what they have seen and integrate it into broader knowledge structures. Emerging findings suggest that such temporal dynamics can sustain or even reshape truth judgments weeks after exposure (Pataranutaporn et al., 2024), underscoring the need to examine how audience perceptions of truth and memory are modified over time.

2.2. Modality and Cognitive Load in Processing Synthetic Narratives: Are Deepfakes Really More Powerful at Distorting Memory Than Text?

Although public debate frequently portrays deepfakes as uniquely dangerous tools capable of deceiving and emotionally manipulating audiences more effectively than traditional media (Weikmann & Lecheler, 2023), empirical evidence paints a more nuanced picture. Experimental research shows that synthetic videos can indeed distort memory, but their effects are not consistently stronger than those of written or photographic narratives of the same events (Murphy & Flynn, 2022). Recent reviews conclude that claims about the exceptional persuasive power of deepfakes are often misguided and that their effects are generally modest and context dependent (Birrer & Just, 2024; Wittenberg et al., 2021).

Beyond technological novelty, the impact of deepfakes depends on how modality shapes information processing. Message format (textual vs. audiovisual) influences the balance between analytic and affective engagement. According to the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986), audiovisual messages usually prompt peripheral processing guided by vivid sensory cues, whereas text invites central, reflective elaboration. The Limited Capacity Model of Motivated Mediated Message Processing (LC4MP; Lang, 2000) and Cognitive Load Theory (CLT; Sweller et al., 2011) further explain that rich multimodal inputs, though immersive, may increase cognitive demands and limit detailed encoding.

Empirical research supports these dynamics. Audiovisual realism enhances emotional transportation and presence (Green & Brock, 2000;

Slater, 2018) but does not necessarily improve truth evaluation or recall accuracy. People often find such messages plausible because of processing fluency rather than analytical reasoning (Fazio et al., 2015). Indeed, comparative experiments reveal that textual formats often foster deeper semantic processing and more stable memory traces than video (Riggenbach et al., 2024). In educational settings, multimodal engagement tends to heighten interest and emotional involvement but this does not necessarily lead to stronger retention (Brame, 2016).

When AI-generated content is clearly labeled as such, viewers are typically more skeptical but still engage emotionally with the material (Gamage et al., 2025). Studies of benign deepfakes of deceased individuals show that even when audiences are aware of their synthetic nature, they often process it as they would a conventional audiovisual narrative—evaluating it more in terms of storytelling quality and emotional resonance than authenticity (Soto-Sanfiel & Wu, 2026). Since fictional narratives are able to promote perceptual fluency and narrative transportation, and therefore immediacy and plausibility (Green & Brock, 2000; Lang, 2000), they can “feel” true even when the viewer knows they are not (Fazio et al., 2015; Newman & Zhang, 2021). This view is supported by evidence from virtual-reality studies suggesting that perceptual realism can heighten emotional and mnemonic responses even when participants are fully aware that the scenario is a simulation (Slater, 2018). Thus, the persuasive potential of deepfakes may stem less from deception than from the sensory and affective qualities of the medium itself.

From a communication perspective, the key question is not simply whether deepfakes can deceive, but whether their audiovisual nature inherently produces stronger cognitive effects than equivalent written narratives. While visual realism may increase immediacy and emotional resonance, text tends to produce greater cognitive elaboration and longer-term retention. Existing findings suggest that modality determines not only how audiences experience synthetic narratives but also how they encode and reassess them over time. When viewers are aware of a deepfake's artificial nature, its realism may heighten analytical vigilance, leading to lower perceived truth but distinct recall patterns. Our study addresses gaps in the literature in this regard by directly comparing, under controlled conditions, the long-term effects of deepfakes and texts in terms of perceived truth and recall.

2.3. Individual differences in cognitive motivation for trusting and recalling deepfakes

Beyond message features, individual differences also shape audience responses to synthetic media. People vary in their motivation to think critically and in their reliance on surface versus analytic processing. These cognitive dispositions may condition memorability and how the truth of deepfakes is judged over time.

Need for Cognition (NFC), the tendency to engage in and enjoy effortful thinking, has long been associated with deeper message scrutiny, improved recall, and reduced reliance on superficial cues such as vividness or modality (Cacioppo & Petty, 1984). NFC moderates media skepticism and trust (Tsfiati & Cappella, 2005) and can reduce susceptibility to the illusory truth effect by encouraging more analytic evaluation of fluent or repeated statements (Newman & Zhang, 2021). Misinformation research finds that individuals who reflect analytically are less likely to believe false claims, endorse fabricated headlines, or share misleading content, even when the material is emotionally charged or stylistically fluent (Bronstein et al., 2019; Martel et al., 2020; Pennycook & Rand, 2019; Ross et al., 2021). These findings show that analytic thinking—and not ideological bias or motivation—best predicts resistance to misinformation across textual and visual formats. The implication is that NFC shapes how audiences interpret and remember synthetic narratives. Individuals high in NFC tend to process information more analytically, whereas those lower in NFC may respond more readily to perceptual fluency and emotional immediacy.

Building on dual-process theory, TDT (Levine, 2021) frames NFC as a

disposition that makes the activation of suspicion more likely, counteracting the truth-default. Comparable patterns emerge in synthetic media contexts: deliberative processors express greater skepticism toward realistic AI-generated content than viewers who rely more on intuition or familiarity (Groh et al., 2024). Cognitive motivation therefore functions as an epistemic brake, tempering the impulse to “believe what looks real.”

However, analytic vigilance chiefly affects truth judgments rather than memory. The illusory truth effect shows that repeated or fluent statements can feel true and later be recalled as such even when recognized as false (Fazio et al., 2015; Hasher et al., 1977). Research on misinformation correction likewise demonstrates that people may correctly identify a claim as false yet still encode and retrieve its details (De Keersmaecker and Roets, 2017; Lewandowsky et al., 2017). Reflective judgment and memory thus rely on partly distinct systems: the former metacognitive and effortful, the latter shaped by fluency, salience, and repetition. This dissociation is particularly relevant to deepfakes. Audiovisual material tends to be vivid, emotionally engaging, and perceptually fluent, making it easy to encode and subsequently retrieve even when its veracity is doubted (Groh et al., 2024). According to the LC4MP (Lang, 2000) and CLT (Sweller et al., 2011), such multimodal input taxes cognitive resources, limiting the capacity for both critical evaluation and detailed rehearsal. Consequently, the trait that supports analytic scrutiny (NFC) may enhance skepticism without improving memory accuracy.

3. Research questions and hypotheses

Drawing on the theoretical and empirical review above, the present study is guided by two research questions (RQ) and two hypotheses (H) concerning how modality and individual cognitive dispositions shape the perception and memory of synthetic historical narratives over time.

RQ1: How does the perceived truth of deepfake narratives evolve over time?

RQ2: How does the recall of deepfake narratives evolve over time?

H1. Modality will influence perceived truth and recall of historical narratives, such that written accounts will elicit stronger truth judgments and recall accuracy than their deepfake video counterparts.

H2. Individuals with higher NFC will attribute lower perceived truth to deepfake narratives than those with lower NFC, but their recall accuracy will not differ significantly.

These hypotheses form a multilevel framework for understanding how audiences perceive, evaluate, and remember synthetic portrayals, in our case of well-known figures. While initial acceptance of deepfakes suggests automatic truth-default processes, the durability of these impressions (RQ1 and RQ2) might depend on factors including modality, as written and audiovisual formats are expected to elicit different balances of analytic and affective engagement (H1). Finally, cognitive motivation determines how critically—but not how accurately—people process synthetic portrayals (H2).

4. Method

4.1. Study design

A longitudinal online experiment conducted via Prolific examined how audiences recall and evaluate the truth of narratives about historical figures presented as text or deepfakes. Using a 2 (protagonist: Albert Einstein vs. Marie Curie) $\times$ 2 (modality: video vs. text) $\times$ 2 (time of exposure: first vs. second message) mixed design across two sessions held one week apart, the study tested how modality, protagonist, and repeated exposure affect memory and perceived truth, controlling for

Need for Cognition and age.

The study was approved by the Ethics Review Board of [University name] and conducted in compliance with the Declaration of Helsinki and relevant institutional guidelines for research involving human participants.

4.2. Participants

A total of 800 participants from the United States were recruited through Prolific. After excluding incomplete responses and failed attention checks, the final sample comprised 332 participants ($M_{age} = 37.02$, $SD = 11.97$, range = 18–79 years). Following the age brackets used in large-scale surveys of public opinion and digital media use, such the report of American Social Media Use released by Pew Research Center (Gottfried, 2024), the sample was further categorized into four age groups: 18–29 ($n = 105$, 31.6%), 30–49 ($n = 169$, 50.9%), 50–64 ($n = 51$, 15.4%), and 65+ ($n = 7$, 2.1%). In terms of gender, 56.0% identified as women, 40.7% as men, 2.7% as non-binary and 0.6% preferred not to say. Most of the participants held a bachelor's degree (47.3%), followed by a master's degree (16.0%). Full demographic characteristics of the sample are reported in Table 1. Participants received USD 2.50 for the first session and USD 3.00 for the second.

4.3. Stimuli

Four first-person biographical narratives were created, each presented either as text or a deepfake video. The stimuli portrayed Albert Einstein and Marie Curie recounting key moments from their scientific and personal lives from a first-person perspective, allowing participants

to engage directly with each protagonist's voice and experiences. Each narrative had two parts (Session 1 and Session 2) of approximately equal length (about 2 min) and comparable content complexity. Their stories traced a progression from the beginnings of their careers to the achievement of scientific acclaim.

All narratives were based on verified biographical information drawn from archival sources and written in a neutral, credible tone. Researchers constructed the narratives using information synthesized from several biographies of the scientists. The scripts for both versions were identical in wording and were introduced as follows: “This message adopts Albert Einstein/Marie Curie's perspective to narrate events or circumstances inspired by his/her life.” Technical details and descriptions of the content are included in Appendix A.

4.4. Measures

4.4.1. Need for Cognition (NFC)

Need for Cognition was measured using the six-item scale developed and validated by Coelho et al. (2020), which captures the stable tendency to engage in and enjoy effortful cognitive activity. Participants indicated the extent to which each statement described them on a five-point scale (1 = does not describe me; 5 = describes me extremely well). Items included statements such as “I prefer complex to simple problems” and “I like to have the responsibility of handling a situation that requires a lot of thinking.” The full item set is available in the open-access questionnaire copy. The six-item version was selected over the original 34-item (Cacioppo & Petty, 1982) or the shortened 18-item version (Cacioppo et al., 1984) because it retains strong psychometric properties while reducing participant burden in the context of a multi-session online experiment (Coelho et al., 2020). Higher scores indicate a greater intrinsic motivation to engage analytically with information. The scale demonstrated good internal consistency in the present sample ($\alpha = .85$).

4.4.2. Recall measurement

Immediately after viewing each narrative (first and second session), participants completed a recognition-style recall test in which they indicated whether fifteen statements about the narrative were true (T), false (F), and unrelated/distractors (D). The statements included.

- True items that appeared in the narrative (5), such as “I began working at the Swiss Federal Patent Office in Bern.”
- False but plausible items that did not appear in the narrative (5), such as “I developed the General Theory of Relativity during my time in Bern.”
- Distractor items that were unrelated to the narrative content but plausible in a general life context (5), such as “I regularly volunteered at a local animal shelter, taking care of abandoned pets.”

This method follows established recognition and false-memory research paradigms in which the combination of true, false, and unrelated items provides a nuanced measure of memory accuracy, error types, and reliance on gist versus verbatim traces (Bücker et al., 2022; Cadavid et al., 2021; Liu et al., 2021). For each participant, we computed: (a) the proportion of true items correctly identified (hits), and (b) the proportion of false items incorrectly endorsed as true (false alarms). A composite recall accuracy score was calculated by summing all correct responses yielding an overall recall performance index. This index allowed comparison across modality (video vs. text), protagonist, and session to show the extent to which participants retained accurate details, formed memory distortions, or responded randomly. Delayed recall used the same items to measure participants' recollection of the same content one week later.

Table 1
Demographic information and balance checks across conditions.

Sample				Balance checks across conditions	
	Category	N	%	Test	Statistics
Age	$M = 37.02$, $SD = 11.97$, range = 18–79	332	NA	ANOVA	modality $\times$ protagonist: $F(1, 328) = 0.01$, $p = .91$
Gender	Woman	186	56.0%	Chi-square	modality: $\chi^2(3) = 0.79$, $p = .85$
	Man	135	40.7%		protagonist:
	Non-binary	9	2.7%		$\chi^2(3) = 0.86$, $p = .84$
	Prefer not to say	2	0.6%		modality: $\chi^2(8) = 10.75$, $p = .22$
Edu.	No formal education	0	0.0%	Chi-square	protagonist: $\chi^2(8) = 4.89$, $p = .77$
	Primary school	17	5.1%		
	Lower secondary (fewer than 5 GCSEs/O-Levels A*–C)	5	1.5%		
	Upper secondary (5+ GCSEs/O-Levels A*–C)	8	2.4%		
	Upper secondary (A-levels)	34	10.2%		
	HNC/CertHE/NVQ Level 4	16	4.8%		
	NVQ Level 5/HND/DipHE/Foundation Degree	29	8.7%		
	Bachelor's degree or equivalent	157	47.3%		
	Master's degree or equivalent	53	16.0%		
	Doctorate or equivalent	13	3.9%		

Note. Balance checks test whether randomization produced demographically equivalent groups. Age tested with ANOVA (modality $\times$ protagonist). Gender and education tested with chi-square separately for modality and protagonist. All tests were non-significant, confirming demographic equivalence across conditions.

4.5. Perceived truth measurement

Participants were then asked to evaluate the extent to which each of the previous statements reflected the real lives of the protagonists on a five-point scale (1 = Does not reflect the real life of the protagonist at all; 5 = Definitely reflects the real life of the protagonist). All items referred to biographically verifiable information about the two historical figures (e.g., “Albert Einstein started his career at the Swiss Federal Patent Office in Bern”). This measure captures participants’ belief in the factual accuracy of narrative and is widely used in research on truth judgments, belief accuracy, and illusory truth effects—the tendency for familiarity and fluency to inflate perceived accuracy even when information is false (e.g., Brashier & Marsh, 2020; Fazio et al., 2015; Pennycook et al., 2018). Responses were averaged to form a composite perceived truth score, with higher values indicating stronger belief in the statements’ factual accuracy. Delayed truth perception measures reused the same content but the items were reworded to measure perceived truthfulness of the earlier message after one week.

4.6. Procedure

After providing demographic information (e.g., gender, age) and completing the NFC scale, participants were informed that they would be randomly assigned to read or watch a first-person narrative of a historical figure presented either as text or video. They were instructed to pay close attention and try to remember as many details as possible. After exposure to each narrative, they completed the immediate recall and perceived truth measures to assess their comprehension and evaluation.

At the one-week follow-up, participants read a second text or watched a second video featuring the same protagonist (Einstein or Curie) as in the first session, who continued narrating subsequent episodes of their life experience. They again completed immediate recall and truth perception measures relating to the second narrative, before being asked to complete delayed recall and truth perception measures about the one they had read or watched a week earlier. This second test assesses how new, related information influences memory and evaluation of earlier material. Finally, participants completed an overall measure evaluating the perceived truthfulness of both messages combined. At the end of the second session, participants were debriefed and informed about the true purpose of the study. They were assured that, although the scientists’ words were fictional, the narratives were based on verified biographical events.

The data and full questionnaires that support the findings of this study are openly available at [https://osf.io/wknfc/overview?view_only=7fe39c21835c4a629faf144825e47f20]. The audiovisual stimuli consist of synthetic reconstructions of deceased historical figures delivering statements grounded in historically documented ideas, although not direct quotations. Because these materials involve performative simulation of identifiable individuals, the files are not publicly archived in order to prevent decontextualized redistribution or misattribution. The deepfake stimuli may be made available to qualified academic researchers upon reasonable request for non-commercial scholarly purposes and under appropriate use conditions. Detailed production procedures and technical specifications are provided to ensure methodological transparency. The text versions of the stimuli are available in Appendix A.

5. Results

Prior to the main analyses, demographic distribution across conditions was checked. No significant differences were found for age, gender, or education, confirming that randomization produced equivalent groups across all conditions (for details see Table 1). To determine whether demographic characteristics warranted statistical control in the main analyses, correlations and one-way ANOVAs were conducted to

examine associations between age, gender, and education and the outcome variables. Age was significantly correlated with immediate recall of the second message, $r = 0.20$, $p < .001$, immediate perceived truth of the first message, $r = 0.18$, $p = .001$, and delayed perceived truth of the first message, $r = 0.15$, $p = .005$. No significant associations were found between age and the remaining outcome measures. Gender and education showed no significant effects on any outcome variable. Age was therefore retained as a covariate in subsequent analyses, while allowing for the possibility that its effects may not be strictly linear.

5.1. Longitudinal effects: perceived truth of the first message

A $2 \times 2 \times 2$ mixed ANOVA examined the effects of modality (text vs. video), protagonist (Einstein vs. Curie), and time (first vs. delayed session) on participants’ perceived truth of Message 1. The analysis revealed significant main effects for all three factors. Text-based narratives elicited higher perceived truth than deepfakes: $F(1, 328) = 13.21$, $p < .001$, $\eta^2 = 0.029$. Einstein’s story was judged more truthful than Curie’s: $F(1, 328) = 15.07$, $p < .001$, $\eta^2 = 0.033$. Addressing RQ1, perceived truth significantly declined over time: $F(1, 328) = 13.67$, $p < .001$, $\eta^2 = 0.011$. Although all main effects were significant, none of the interactions were significant. In line with H1, written narratives were perceived as more truthful than deepfakes, both initially and after a one-week delay (Fig. 1).

5.2. Longitudinal effects: recall accuracy of the first message

A $2 \times 2 \times 2$ mixed ANOVA examined the effects of modality, protagonist, and time on participants’ recall of the first message. All three factors presented significant main effects. Recall was significantly higher for text than for deepfakes: $F(1, 328) = 20.95$, $p < .001$, $\eta^2 = 0.042$, and higher for Curie than for Einstein: $F(1, 328) = 34.22$, $p < .001$, $\eta^2 = 0.067$. Addressing RQ2, recall accuracy significantly declined over time: $F(1, 328) = 9.86$, $p = .002$, $\eta^2 = 0.009$. Of the interaction effects, only modality $\times$ time reached significance: $F(1, 328) = 4.02$, $p = .046$, $\eta^2 = 0.004$. Recall decreased more sharply for deepfake videos than for written accounts. Supporting H1, written narratives elicited more accurate and durable recall than their deepfake counterparts, suggesting that text-based formats foster stronger long-term retention, while audiovisual presentations are more prone to memory decay over time (Fig. 2).

5.3. Multivariate effects of modality, protagonist and time on recall and perceived truth

A multivariate analysis of covariance (MANCOVA) examined the combined effects of protagonist and modality on seven dependent variables: immediate recall of the first message (Session 1), immediate recall of the second message (Session 2), delayed recall of the first message (Session 2), immediate perceived truth of the first message (Time 1), immediate perceived truth of the second message (Session 2), delayed perceived truth of the first message (Time 2), and overall perceived truth of both messages (Time 2). NFC and age were included as covariates. Significant multivariate main effects emerged for protagonist, Pillai’s Trace = 0.21, $F(7, 320) = 12.29$, $p < .001$; modality, Pillai’s Trace = 0.11, $F(7, 320) = 5.90$, $p < .001$; NFC, Pillai’s Trace = 0.08, $F(7, 320) = 3.84$, $p < .001$; and age, Pillai’s Trace = 0.09, $F(7, 320) = 4.26$, $p < .001$. The interaction between protagonist and modality was marginally significant, Pillai’s Trace = 0.04, $F(7, 320) = 2.02$, $p = .052$. Follow-up univariate ANCOVAs identified which specific measures were affected by the protagonist, modality, and their interaction, while controlling for NFC and age (See Fig. 3). A Bonferroni-adjusted significance threshold ($\alpha = .0071$) was applied to control for multiple comparisons.

For immediate recall of the first message, there were significant main effects of protagonist, $F(1, 326) = 23.75$, $p < .001$, and presentation format, $F(1, 326) = 20.95$, $p < .001$; both remained significant after

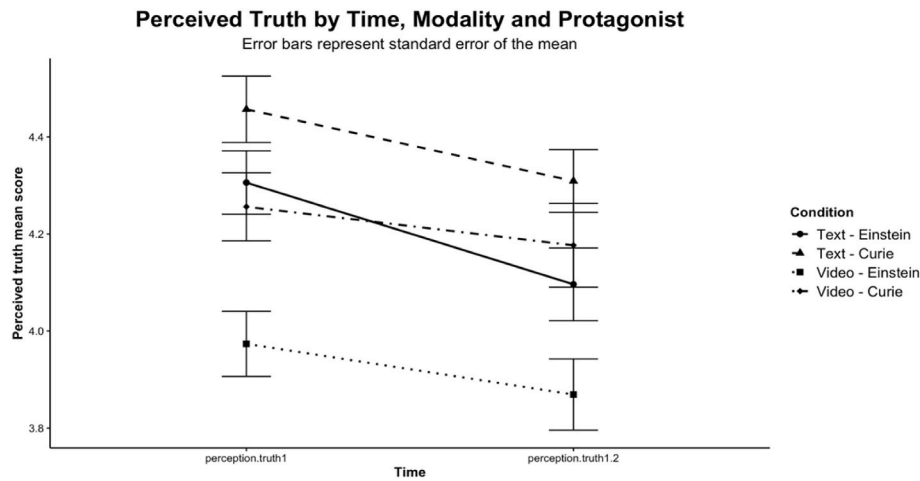


Fig. 1. Difference in means of perceived truth across conditions.

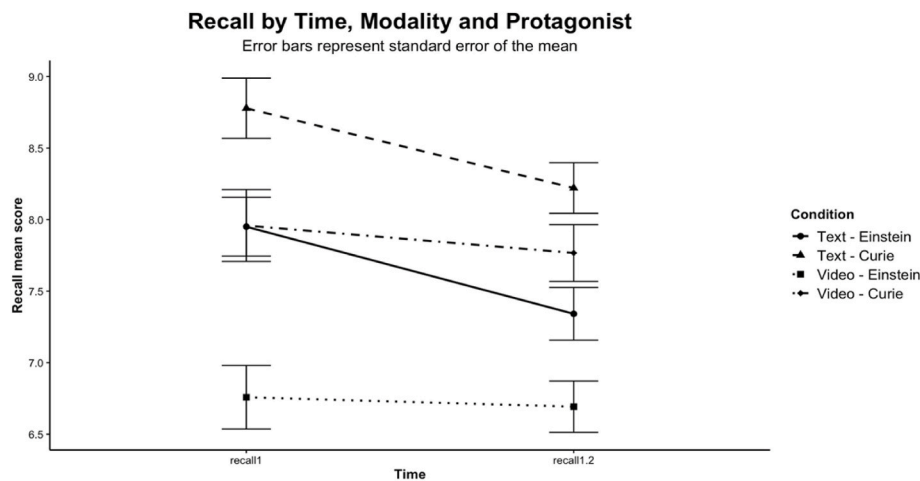


Fig. 2. Difference in means of recall of the first message across conditions.

correction. For immediate recall of the second message, strong main effects emerged for protagonist, $F(1, 326) = 42.50, p < .001$, modality, $F(1, 326) = 18.39, p < .001$, and age, $F(1, 326) = 15.65, p < .001$, all of which remained significant after correction. A protagonist $\times$ modality interaction also reached significance, $F(1, 326) = 12.10, p < .001$, showing that the advantage of text over deepfake was particularly pronounced for Einstein's narrative. For delayed recall of the first message, both protagonist, $F(1, 326) = 30.52, p < .001$, and modality, $F(1, 326) = 9.17, p = .003$, remained significant after correction. No significant effects of NFC or age were observed for recall measures.

For immediate perceived truth of the first message, significant main effects were found for protagonist, $F(1, 326) = 12.76, p < .001$; modality: $F(1, 326) = 16.95, p < .001$; NFC: $F(1, 326) = 15.91, p < .001$; and age: $F(1, 326) = 11.95, p < .001$ —all of which remained significant after Bonferroni correction. Truth judgments were higher for Einstein, for text-based narratives, and among participants with higher NFC or age. For immediate perceived truth of the second message, only NFC had a significant effect after correction, $F(1, 326) = 10.91, p = .001$. For delayed perceived truth of the first message, effects of protagonist ($F(1, 326) = 14.32, p < .001$), NFC ($F(1, 326) = 20.95, p < .001$), and age ($F(1, 326) = 7.71, p = .006$) all remained significant after correction. Persistence in perceived truth was stronger for Einstein's narrative and reflected differences driven primarily by the lowest scores observed among the youngest participants (see analysis on age group below). For overall perceived truth across both messages, significant main effects

were found for protagonist, $F(1, 326) = 17.60, p < .001$ and NFC, $F(1, 326) = 11.54, p < .001$.

Across all measures, the protagonist exerted consistent effects on both recall and perceived truth, while presentation format primarily influenced recall and, to a lesser extent, initial truth judgments. Contrary to the first part of H2, NFC positively predicted truth perception. Consistent with the second part of H2, NFC did not correlate with recall. Age influenced immediate recall of the second message and initial truth ratings, and a single significant interaction between protagonist and format emerged for immediate recall of the second message, indicating condition-specific variation in cognitive engagement.

Since the sample was skewed toward younger and middle-aged adults, we conducted supplementary analyses to further examine the structure of the age effects reported above. MANCOVA was re-estimated with age entered as a four-level categorical factor (18–29, 30–49, 50–64, 65+), while retaining protagonist, modality, NFC, and the protagonist $\times$ modality interaction. The multivariate effect of age group was significant, Pillai's Trace = 0.12, $F(21, 960) = 1.96, p = .006$. The pattern of univariate effects was consistent with the continuous specification where age group significantly predicted immediate recall of the second message, $F(3, 324) = 6.95, p < .001$, and immediate perceived truth of the first message, $F(3, 324) = 3.65, p = .013$, and delayed perceived truth of the first message, $F(3, 324) = 3.97, p = .008$.

Bonferroni-corrected pairwise comparisons based on estimated marginal means showed that the 18–29 group ($M = 6.91$) recalled

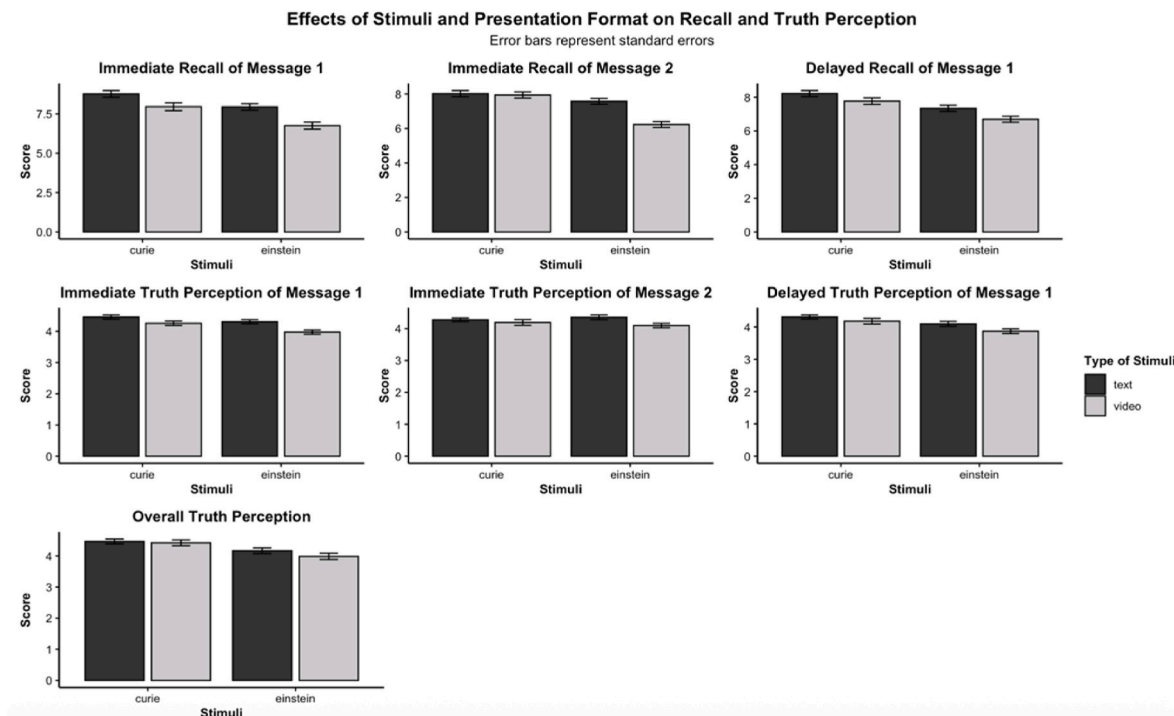


Fig. 3. Effects of modality and protagonist on recall and perceived truth.

significantly fewer details of the second message than the 30–49 group ($M = 7.62, p = .002$) and the 50–64 group ($M = 7.80, p = .006$), with a borderline difference relative to the 65+ group ($M = 8.44, p = .07$). For delayed perceived truth of the first message, participants aged 18–29 ($M = 3.95$) scored significantly lower than those aged 30–49 ($M = 4.19, p = .018$). No other age-group pairs differed significantly on any measure after Bonferroni correction.

The convergence between continuous and categorical specifications increases confidence in the robustness of the observed age effects. Importantly, supplementary analyses showed that age effects were not driven by the small number of participants aged 65+, as no significant differences were observed between the older age groups. Rather than substantiating a gradual increase in truth perception and recall across age, the results suggest a discontinuity in which the youngest cohort differs from all other groups. In other words, the observed age effect reflects a relative disadvantage of the youngest cohort, which consistently scores lower than participants aged 30 and above.

6. Discussion

This study has examined how modality, protagonist, and cognitive motivation shape the recall and perceived truth of historical narratives presented as text or deepfake video across two time points. Overall, the results partially support our hypotheses: while both recall and perceived truth declined over time, written accounts consistently produced higher recall accuracy and stronger perceptions of truth than their deepfake counterparts, contradicting the expectation that audiovisual realism would enhance credibility.

Perceived truth and recall decreased over time across conditions, suggesting a natural decay of belief strength as memory traces fade. This finding aligns with TDT (Levine, 2014, 2021), which posits that people initially accept information as true unless prompted by cues of potential falsity. In our experiment, participants' initial judgments of truth were moderately high across modalities, reflecting a truth-default orientation during first exposure. However, after one week, the reduction in perceived truth suggests that temporal distance weakens this default acceptance, especially when the message is not reinforced, which is a

pattern consistent with research showing that memory decay and reduced familiarity can diminish belief certainty (Fazio et al., 2015).

Supporting H1, initial perceived truth was higher for text-based narratives than for deepfakes. Rather than boosting credibility, audiovisual realism appeared to trigger evaluative scrutiny. Research on misinformation labeling and synthetic media suggests that awareness of manipulation fosters corrective vigilance and skepticism (Hameleers and Marquart, 2023). From a TDT perspective, the deepfake format likely increased perceived risk of deception, prompting a shift from truth-default to more analytical evaluation. This finding extends Groh et al. (2024), who observed that synthetic political videos elicit greater caution and lower accuracy ratings than textual transcripts of the same content.

Findings from both sessions also strongly support modality-based differences in recall (H1). Participants who read the text-based narratives remembered significantly more details than those who viewed videos, both immediately and after one week. This pattern underscores how text fosters deeper cognitive processing, consistent with the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) and the Limited Capacity Model of Motivated Mediated Message Processing (LC4MP; Lang, 2000). Written narratives require sustained attention and elaboration, promoting central-route encoding, whereas audiovisual messages impose higher cognitive load and distribute attention across multiple sensory channels, which can hinder detailed encoding of information (Sweller et al., 2011; Lang, 2000).

The reduced recall for deepfake stimuli also aligns with Cognitive Load Theory (CLT), whereby multimodal messages require parallel processing of visual, auditory, and semantic cues, increasing extraneous load and limiting resources for schema construction and long-term storage (Sweller, 2011). Similar effects have been documented in studies comparing audiovisual and textual persuasion, where videos heighten emotional engagement but weaken analytical recall (Wittenberg et al., 2021). Thus, the sensory vividness of deepfakes may enhance immersion but simultaneously impair detailed retention, echoing findings that realistic formats promote gist-based rather than verbatim memory (Reyna et al., 2016).

While controlling for individual differences, we found a significant

protagonist $\times$ modality interaction in the recall performance of the second message where the text-over-deepfake advantage was particularly pronounced for Einstein's narrative compared to Curie's. This interaction invites interpretation beyond a simple modality effect. Einstein is arguably one of the most globally recognizable scientific figures, with a comparatively more familiar biographical profile that most participants are likely to have encountered prior to the study. When presented as text, narratives that align with pre-existing schemas are processed through a central route (Petty & Cacioppo, 1986), with participants actively integrating new details into well-established knowledge structures, which is a process that facilitates verbatim encoding and subsequent recall. When the same familiar narrative is presented as a deepfake, however, the audiovisual format may have produced cognitive conflict where participants must simultaneously process the biographical content, the synthetic visual cues, and the incongruence between Einstein's iconic image and its AI-generated rendering. This incongruence may increase extraneous cognitive load (Sweller et al., 2011), diverting resources away from detailed encoding and thereby widening the modality gap in recall relative to a less iconically burdened figure such as Curie. For Curie, where prior schema activation is comparatively weaker, the deepfake format may not trigger the same level of conflict, moderating the differential. Therefore, the cognitive effects of presentation modality may not be uniform across content but depend on the interaction between format and audience familiarity with the portrayed figure. Future research should verify these assumptions.

With regards to truth perception, despite earlier research reporting that visual realism can strengthen emotional and mnemonic responses (Gamage et al., 2025; Slater, 2018), participants in our study judged text narratives as more truthful than deepfakes. This divergence parallels findings in cognitive psychology that perceptual fluency does not uniformly translate into credibility (Unkelbach & Rom, 2017). In the context of deepfakes, realism may instead increase epistemic tension: participants recognize lifelike audiovisual cues despite being aware of potential artificiality. The perceived truth judgments in our study align with the illusory truth effect, whereby repeated or fluent information feels accurate even when it is objectively false (Fazio et al., 2015; Pennycook et al., 2018). The mechanism underlying this effect has been described as "referential consolidation", whereby repetition strengthens associative networks around a claim, making it feel corroborated and therefore "truer" (Unkelbach & Rom, 2017; Brashier & Marsh, 2020).

Our analyses also showed that NFC significantly predicts perceived truth, but in the opposite direction to that proposed by H2. Individuals high in NFC—those who enjoy and engage in effortful thinking—tended to rate narratives as more truthful overall. Although this may seem counterintuitive, dual-process frameworks maintain that high-NFC individuals rely less on superficial cues and evaluate content through systematic reasoning (Cacioppo & Petty, 1984; Pennycook & Rand, 2019), typically making them more skeptical and less likely to trust information uncritically. However, because all perceived truth items used in our study reflected accurate biographical statements, high-NFC participants likely processed them more attentively, yielding more accurate judgments of truth. From an ELM perspective, these participants were more likely to follow the central processing route, integrating message content with prior knowledge and semantic coherence, leading to more positive and stable evaluations. Meanwhile, the absence of a comparable NFC effect on recall suggests that cognitive motivation influences evaluative judgment more than memory storage. Recall primarily reflects perceptual and encoding efficiency rather than epistemic evaluation, while truth judgments involve higher-order reasoning about plausibility. This dissociation mirrors findings by Fazio et al. (2015) and Brashier and Marsh (2020), who demonstrated that belief formation and factual recall operate through distinct cognitive pathways, with the former more responsive to analytic dispositions and critical reasoning skills.

Beyond individual differences in cognitive motivation, age emerged as a significant covariate across several dependent measures. However,

closer inspection of the data indicates that this effect does not reflect a linear increase across age groups. Instead, it is best understood as a cohort-based discontinuity driven by the youngest participants. Supplementary analyses using age as a categorical factor clarify that participants aged 18–29 consistently scored lower than those aged 30 and above on immediate recall of the second message, as well as on delayed perceived truth of the first message. Importantly, no significant differences were observed among the groups aged 30–49, 50–64, and 65+, indicating that the effect is not attributable to a progressive advantage among older participants but rather to a relative disadvantage of the youngest cohort. One plausible explanation for this pattern lies in age-related differences in preferred processing orientations when engaging with deepfake stimuli, alongside variation in familiarity with the depicted historical figures (Tang et al., 2025). It is possible that participants aged 30 and above are more likely to possess consolidated knowledge structures related to historical figures such as Einstein and Curie, which may facilitate both encoding and evaluation of the narratives. When information aligns with existing schemas, it is processed more fluently and integrated more effectively into memory, resulting in stronger and more stable truth judgments (Brashier & Marsh, 2020). By contrast, younger participants may rely less on established knowledge structures and more on surface-level or format-related cues when processing the stimuli (Tang et al., 2025). In the case of deepfakes, this may increase cognitive load and divert attention away from semantic encoding, particularly when prior familiarity with the depicted figures is limited. This interpretation is consistent with research showing that audiovisual complexity can interfere with deep processing when cognitive resources are distributed across multiple channels (Lang, 2000; Sweller et al., 2011).

Importantly, the age effects uncovered in this study should not be interpreted as indicating a general advantage of older adults. Rather, they highlight the role of familiarity and prior knowledge in shaping both memory and belief. The observed differences reflect the position of the youngest cohort relative to the rest of the sample, rather than a monotonic age-related trend. These conclusions should therefore be interpreted with caution, given the age distribution of the sample, which was skewed toward younger and middle-aged adults and included a limited number of participants over 65. Future research with more balanced age distributions, particularly with larger samples of older adults, is needed to confirm whether these patterns hold across the full adult lifespan and to disentangle schema-based explanations from cohort-specific differences in media experience.

Taken together, our findings illustrate a complex relationship between modality, cognition, and truth perception. Written accounts fostered both stronger memory and greater perceived truth, whereas deepfakes were processed more superficially and remembered less accurately. Over time, perceived truth declined across modalities, suggesting that memory decay weakens belief certainty even in non-deceptive contexts. The persistence of protagonist effects further suggests that familiarity and cultural salience shape both memory and belief. Participants may have found Einstein's story more familiar, enhancing perceptual fluency, while Curie's narrative demanded greater cognitive reconstruction, leading to lower recall. Einstein remains a globally iconic figure, while Curie, despite her scientific renown, is less familiar to mainstream audiences. These results challenge deterministic assumptions about technological realism. Although deepfakes are often presumed to enhance credibility through vividness and immersion, this study suggests the opposite: realism may cause interpretive resistance, prompting critical evaluation rather than naïve acceptance. Empirically, the findings demonstrate that audiovisual realism alone does not guarantee belief. Instead, trust in mediated information depends on other factors, such as cognitive motivation and interpretive context.

As the findings emphasize that fostering analytical engagement may counteract susceptibility to misinformation, even when messages are visually persuasive, they have clear implications for AI literacy education and media ethics. But they also speak to a broader cultural concern.

Synthetic reconstructions of the past may gradually reshape collective memory and public understanding of history. Scholars have warned that deepfakes and other AI-generated media can erode epistemic trust in authentic records and encourage subtle forms of historical revisionism (Pataranutaporn et al., 2024; American Historical Association, 2021). By clarifying how truth judgments evolve when viewers are aware of a video's artificial nature, this study advances both theoretical and practical insight into how realism, modality, and cognition interact in a post-truth media environment.

7. Limitations and further research

This study has certain limitations that suggest directions for future research. The one-week interval captured short-term changes in memory and belief but may not reflect longer-term consolidation or updating. Extended longitudinal designs could clarify how perceived truth evolves over time. The use of benign historical figures limits generalization to emotionally or politically charged deepfakes, where affect and ideology may interact with modality to shape recall and credibility. Although message length and structure were controlled, text and video inherently differ in cognitive and perceptual demands, making it difficult to isolate modality effects from overall cognitive load. In addition, the U.S.-based online sample may not be representative of audiences from different cultural backgrounds or with varying levels of AI literacy. Finally, reliance on recognition and self-report measures may underestimate implicit belief formation. Future research should therefore incorporate cross-cultural samples, emotionally salient stimuli, and behavioral or physiological measures to deepen understanding of how modality and cognition jointly shape memory and truth judgments in deepfake contexts.

8. Conclusion

Despite the widespread concern about the potential for deepfakes to distort history and erode public trust, this study reveals that written historical narratives elicit both stronger memory and higher perceived truth than their deepfake counterparts. These effects persisted despite the sensory realism of deepfakes, indicating that technological vividness does not equate to credibility. Instead, cognitive engagement, particularly among individuals high in Need for Cognition, emerged as the most consistent predictor of how audiences evaluate and remember mediated representations of the past. Rather than revealing mass vulnerability, the findings highlight how critical thinking and media awareness can serve as protective factors against manipulation. Nonetheless, the cultural and emotional power of realism in digital media continues to challenge how societies negotiate authenticity, historical memory, and the boundary between representation and fabrication. Future research should extend this work to emotionally charged or politically framed deepfakes, where affective arousal may interact with cognitive load and fluency in other ways. Furthermore, cross-cultural comparisons could clarify whether interpretive vigilance toward synthetic media varies across epistemic cultures, while more longitudinal designs could trace how belief updating and memory distortion evolve over time, enriching understanding of how synthetic realism interacts with collective memory.

CRedit authorship contribution statement

María T. Soto-Sanfiel: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.
Gina Junhan Fu: Writing – review & editing, Writing – original draft, Investigation, Formal analysis.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) exclusively to assist with grammatical refinement, as the authors are not native speakers of English. After using this tool, the authors thoroughly reviewed and edited the text, and they take full responsibility for the content and accuracy of the published article.

Funding

This research has received support from the Centre for Trusted Internet and Community and the Faculty of Arts and Social Sciences, National University of Singapore (CTIC-CP-21-07/A-8001097-00-00). María T. Soto-Sanfiel, grant holder.

Declaration of competing interest

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

Acknowledgment

The authors wish to thank Sanjay Saha for his valuable collaboration in the production of the experimental stimuli.

Appendix A

Technical Details and Textual Content of the Narratives by Historical Figure and Session.

The deepfake videos combined synthetic visuals with cloned voices. For the visuals, a Thin Plate Spline motion model was applied for face reenactment. Archival portraits of Einstein and Curie were selected and edited, and two professional actors (one male, one female) were filmed delivering the monologues directly to the camera. Their facial and upper-body movements were then mapped onto the static portraits to produce lifelike animations. For the audio, the ElevenLabs voice-cloning tool was used. Cleaned and normalized archive recordings of Einstein and Curie served as source material to train synthetic voice models capable of generating new speech synchronized with the actors' lip movements. After post-processing, visuals and audio were merged to create the complete deepfake videos.

Albert Einstein

Session 1

When I began working at the Swiss Federal Patent Office in Bern, I found myself in a position quite different from the academic circles I had aspired to join. Nonetheless, it was during this time that I would later come to call my “miracle year.” The quiet office, filled with technical documents and patent applications, provided me with ample time to ponder the mysteries of the universe.

I have often said, “A happy man is too satisfied with the present to dwell too much on the future.” In Bern, I was content, and my thoughts could freely wander. It was here that I developed some of my most significant theories, including the Special Theory of Relativity. I would spend hours contemplating the nature of time and space, envisioning what it would be like to ride alongside a beam of light.

In those days, I also formed a group with my close friends, the “Olympia Academy,” where we would gather to discuss scientific ideas, philosophy, and literature. These spirited conversations fueled my intellect and imagination, pushing me to delve deeper into the unknown.

Session 2

During that miraculous year, I published four groundbreaking papers that tackled the most pressing scientific questions of the time. The first, on the photoelectric effect, stemmed from late-night conversations with my wife Mileva, as we mulled over Max Planck's work on blackbody radiation.

The second and third papers, exploring atoms and their motion, were inspired by a simple question from my son: "Papa, why does the dust in the air move?" This childlike curiosity led me to investigate Brownian motion and support the reality of atoms and molecules.

My fourth paper, the Special Theory of Relativity, was a monumental leap that forever changed our understanding of the universe. This theory was born from countless discussions and debates with my close friends in the "Olympia Academy."

Marie Curie

Session 1

As I look back on my journey, I am reminded of the vital role the "Floating University" played in the early stages of my scientific awakening. In those days, my homeland of Poland was under Russian rule, and women were denied access to higher education. The thirst for knowledge, however, could not be quenched by the oppressive regime. It was amidst this climate of intellectual rebellion that the Floating University came to be.

The Floating University was an underground network of informal lectures and classes, held in secret locations throughout Warsaw. It was a beacon of hope, a sanctuary for those of us who longed to learn and explore the world of science, literature, and philosophy. I would attend these clandestine gatherings with my sister Bronisława, sharing in the exhilaration of our collective pursuit of knowledge.

During my time at the Floating University, I was exposed to a variety of subjects, but it was the study of physics and chemistry that ignited a spark within me. I found myself captivated by the underlying principles that governed the natural world, and I knew that I wanted to dedicate my life to unraveling its mysteries.

Session 2

My heart swells with pride and awe as I recount the momentous discovery of two new radioactive elements: polonium, named in honor of my beloved homeland, Poland, and radium, whose very name evokes its radiant properties.

Working alongside my dear husband, Pierre, we embarked on a quest to unravel the mysterious phenomenon of radioactivity. In a cramped, poorly ventilated laboratory, we spent countless hours hunched over our equipment, meticulously analyzing the mineral pitchblende, which we believed held the key to unlocking the secrets of this enigmatic force.

The painstaking process of extracting and purifying these elusive elements demanded immense patience and precision. Yet, in the face of these obstacles, our determination only grew stronger.

Data availability

Data is available at an open source repository

References

- Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, 7, Article 1400024. <https://doi.org/10.3389/fdata.2024.1400024>
- American Historical Association. (2021). *Keeping it real: Historians in the deepfake era. Perspectives on History*. <https://www.historians.org/perspectives-article/keeping-it-real-historians-in-the-deepfake-era-may-2021/>.

- Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448241253138>
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2
- Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE-Life Sciences Education*, 15(4). <https://doi.org/10.1187/cbe.16-03-0125.es6>
- Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, 71, 499–515. <https://doi.org/10.1146/annurev-psych-010419-050807>
- Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, 8(1), 108–117. <https://doi.org/10.1037/h0101832>
- Bücken, C. A., Otgaar, H., Sauerland, M., & Howe, M. L. (2022). Simulating denial increases false memory rates for abuse-related narratives. *Behavioural Sciences & the Law*, 40(5), 702–716. <https://doi.org/10.1002/bsl.2566>
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cacioppo, J. T., & Petty, R. E. (1984). Need for cognition: A motivational factor in information processing. *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13
- Cadavid, S., Beato, M. S., & Fernández, A. (2021). False memories for emotional and neutral scenes: The role of gist and detail processing. *Frontiers in Psychology*, 12, Article 686390. <https://doi.org/10.3389/fpsyg.2021.686390>
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1819. <https://doi.org/10.15779/Z38RV0D15J>
- Coelho, G. L. de H., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8), 1870–1885. <https://doi.org/10.1177/1073191118793208>
- De Keersmaecker, J., & Roets, A. (2017). "Fake news": Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, 65, 107–110. <https://doi.org/10.1016/j.intell.2017.10.005>
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>
- Deck, S. L., & Paterson, H. M. (2020). Liars are perceived as more credible than truth-tellers who recall a repeated event. *Applied Cognitive Psychology*, 34(3), 643–653. <https://doi.org/10.1002/acp.3649>
- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002. <https://psycnet.apa.org/doi/10.1037/xge0000098>
- Gamege, D., Sewwandi, D., Zhang, M., & Bandara, A. K. (2025). Labeling synthetic content: User perceptions of label designs for AI-generated content on social media. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–29). Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713171>. Article 814.
- Gottfried, J. (2024). Americans' social media use. <https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/>.
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. <https://doi.org/10.1037/0022-3514.79.5.701>
- Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. *Nature Communications*, 15(1), 7629. <https://doi.org/10.1038/s41467-024-51998-z>
- Hameleers, M., & Marquart, F. (2023). It's nothing but a deepfake! the effects of misinformation and deepfake labels delegitimizing an authentic political speech. *International Journal of Communication*, 17, 6291–6311. <https://ijoc.org/index.php/ijoc/article/view/20777/4351>
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
- Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, 50(1), 46–70. <https://doi.org/10.1111/j.1460-2466.2000.tb02833.x>
- Lees, D. (2024). Deepfakes in documentary film production: images of deception in the representation of the real. *Studies in Documentary Film*, 18(2), 108–129. <https://doi.org/10.1080/17503280.2023.2284680>
- Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378–392. <https://doi.org/10.1177/0261927X14535916>
- Levine, T. R. (2021). Distrust, false cues, and below-chance deception detection accuracy: Commentary on Stel et al. (2020) and further reflections on (un)conscious lie detection from the perspective of Truth-Default Theory. *Frontiers in Psychology*, 12, Article 642359. <https://doi.org/10.3389/fpsyg.2021.642359>
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Li, F., & Yang, Y. (2024). Impact of artificial intelligence-generated content labels on perceived accuracy, message credibility, and sharing intentions for misinformation:

- Web-based, randomized, controlled experiment. *JMIR Formative Research*, 8, Article e60024. <https://doi.org/10.2196/60024>
- Liu, Z., Liu, T., & Li, Y. (2021). How does social competition affect true and false recognition? *Psychonomic Bulletin & Review*, 28(1), 292–303. <https://doi.org/10.3758/s13423-020-01807-7>
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>
- Lundberg, E., & Mozelius, P. (2025). The potential effects of deepfakes on news media and entertainment. *AI & Society*, 40(4), 2159–2170. <https://doi.org/10.1007/s00146-024-02072-1>
- Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research*, 49(2), 171–195. <https://doi.org/10.1177/0093650220921321>
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, 5, 47. <https://doi.org/10.1186/s41235-020-00252-3>
- Mihailova, M. (2021). To Dally with Dalf: Deepfake (inter)faces in the art museum. *Convergence*, 27(4), 882–898. <https://doi.org/10.1177/13548565211029401>
- Murillo-Ligorred, V., Pertegal-Vega, M. A., & García-Jiménez, A. (2023). Knowledge, integration and scope of deepfakes in arts and education: Challenges and opportunities for critical media literacy. *Education Sciences*, 13(11), 1073. <https://doi.org/10.3390/educsci13111073>
- Murphy, G., Ching, D., Meehan, E., Twomey, J., Bolger, A., & Linehan, C. (2025). An average Joe, a laptop, and a dream: Assessing the potency of homemade political deepfakes. *Applied Cognitive Psychology*, 39(2), Article e70061.
- Murphy, G., de Saint Laurent, C., Reynolds, M., Aftab, O., Hegarty, K., Sun, Y., & Greene, C. M. (2023). What do we study when we study misinformation? A scoping review of experimental research (2016–2022). *Harvard Kennedy School Misinformation Review*, 4(6). <https://doi.org/10.37016/mr-2020-130>
- Murphy, G., & Flynn, E. (2022). Deepfake false memories. *Memory*, 30(4), 480–492.
- Newman, E. J., & Zhang, L. (2021). Truthiness: How non-probative photos shape belief. In R. Greifeneder, M. E. Jaffé, E. J. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 90–114). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780429295379-8>
- Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, 119(8), Article e2120481119. <https://doi.org/10.1073/pnas.2120481119>
- P'Ng, J. (2024). The resurrection will not be televised: Legal remedies for posthumous deepfakes. *The Georgetown Law Technology Review*, 8(2), 338+. <https://link.gale.com/apps/doc/A797177063/LT?u=nuslib&sid=summon&xid=23f38e9a>
- Pataranutaporn, P., Archiwanguprok, C., Chan, S. W. T., Loftus, E., & Maes, P. (2024). Synthetic human memories: AI-edited images and videos can implant false memories and distort recollection. *arXiv preprint arXiv:2409.08895*. <https://arxiv.org/abs/2409.08895>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7(1), 1–75. [https://doi.org/10.1016/1041-6080\(95\)90031-4](https://doi.org/10.1016/1041-6080(95)90031-4)
- Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, 5(1), 1–9. <https://doi.org/10.1016/j.jarmac.2015.12.003>
- Riggenbach, M. R., Gronlund, S. D., & Zoladz, P. R. (2024). Impact of modality and mode of questioning and testing on memory reports. *Frontiers in Cognition*, 5, Article 1349511. <https://doi.org/10.3389/fcogn.2024.1349511>
- Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, 16(2), 484–504. <https://doi.org/10.1017/S1930297500008640>
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431–433. <https://doi.org/10.1111/bjop.12305>
- Soto-Sanfiel, M. T., Angulo-Brunet, A., & Saha, S. (2025). Deepfakes as narratives: Psychological processes explaining their reception. *Computers in Human Behavior*, 165, Article 108518. <https://doi.org/10.1016/j.chb.2024.108518>
- Soto-Sanfiel, M. T., & Wu, Q. (2026). Audiences' cognitive and affective responses to deepfake resurrections of deceased artists. *Computers in Human Behavior*, 174, Article 108822. <https://doi.org/10.1016/j.chb.2025.108822>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
- Tang, Z., Goh, D. H. L., Lee, C. S., & Chen, C. (2025). Understanding deepfake detection strategies among young adults and seniors. *International Journal of Human-Computer Interaction*, 1–19. <https://doi.org/10.1080/10447318.2025.2607571>
- Tsfati, Y., & Cappella, J. N. (2005). Why do people watch news they do not trust? The need for cognition as a moderator in the association between news media skepticism and exposure. *Media Psychology*, 7(3), 251–271. https://doi.org/10.1207/S1532785XMEP0703_2
- Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, 160, 110–126. <https://doi.org/10.1016/j.cognition.2016.12.016>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1–13. <https://doi.org/10.1177/2056305120903408>
- Weikmann, T., & Lecheler, S. (2023). Cutting through the hype: Understanding the implications of deepfakes for the fact-checking actor-network. *Digital Journalism*, 12(10), 1505–1522. <https://doi.org/10.1080/21670811.2023.2194665>
- Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2021). The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences of the United States of America*, 118(47), Article e2114388118. <https://doi.org/10.1002/acp.70061>