

RESEARCH ARTICLE

The public mental representations of deepfake technology: An in-depth qualitative exploration through Quora text data analysis

Barbara Caci¹, Giulia Giordano^{1*}, Marianna Alesi¹, Ambra Gentile¹, Chiara Agnello¹, Liliana Lo Presti², Marco La Cascia², Sonia Ingoglia¹, Cristiano Inguglia¹, Alice Volpes³, Dario Monzani^{1,4}

1 Department of Psychology, Educational Sciences and Human Movement, University of Palermo, Palermo, Italy, **2** Department of Engineering, University of Palermo, Palermo, Italy, **3** Independent Researcher, Padova, Italy, **4** Applied Research Division for Cognitive and Psychological Science, European Institute of Oncology IRCCS, IEO, Milan, Italy

* giulia.giordano@unipa.it



OPEN ACCESS

Citation: Caci B, Giordano G, Alesi M, Gentile A, Agnello C, Lo Presti L, et al. (2024) The public mental representations of deepfake technology: An in-depth qualitative exploration through Quora text data analysis. PLoS ONE 19(12): e0313605. <https://doi.org/10.1371/journal.pone.0313605>

Editor: Yogan Jaya Kumar, Universiti Teknikal Malaysia Melaka Fakulti Teknologi Maklumat dan Komunikasi, MALAYSIA

Received: May 13, 2024

Accepted: October 3, 2024

Published: December 30, 2024

Copyright: © 2024 Caci et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: We are unable to make our dataset available, due to Quora's terms of service. However, we do provide details of the search parameters that we used to retrieve the data from Quora internal search function. A complete list of URLs of collected Quora's questions were included in the Supplementary file. The European Union General Data Protection Regulation 2016/679 allows for the use of anonymous data for research purposes under certain conditions. Since all analyses have been performed on public and

Abstract

The advent of deepfake technology has raised significant concerns regarding its impact on individuals' cognitive processes and beliefs, considering the pervasive relationships between technology and human cognition. This study delves into the psychological literature surrounding deepfakes, focusing on people's public representation of this emerging technology and highlighting prevailing themes, opinions, and emotions. Under the media framing, the theoretical framework is crucial in shaping individuals' cognitive schemas regarding technology. A qualitative method has been applied to unveil patterns, correlations, and recurring themes of beliefs about the main topic, deepfake, discussed on the forum Quora. The final extracted text corpus consisted of 166 answers to 17 questions. Analysis results highlighted the 20 most prevalent critical lemmas, and deepfake was the main one. Moreover, co-occurrence analysis identified words frequently appearing with the lemma deepfake, including video, create, and artificial intelligence—finally, thematic analysis identified eight main themes within the deepfake corpus. Cognitive processes rely on critical thinking skills in detecting anomalies in fake videos or discerning between the negative and positive impacts of deepfakes from an ethical point of view. Moreover, people adapt their beliefs and mental schemas concerning the representation of technology. Future studies should explore the role of media literacy in helping individuals to identify deepfake content since people may not be familiar with the concept of deepfakes or may not fully understand the negative or positive implications. Increased awareness and understanding of technology can empower individuals to evaluate critically the media related to Artificial Intelligence.

Introduction

Deepfake technology, powered by Artificial Intelligence (AI) techniques associated with advanced machine learning algorithms [1], has ushered in a new era of realistic and

anonymized data, no institutional review board approval was required for the use of this database or the completion of this study

Funding: This research was supported by the project titled: DeepFake e Identità Digitale (Deepfake and Digital Identity) – Bando UNIPA Incentivi ad Attività di Ricerca Interdisciplinare; we applied for the fundings we receive from University of Palermo. UNIPA is the acronym of UNiversity of PALermo. (D.R. 4186/2023; prot. 90342 del 15/6/2023).

Competing interests: The authors have declared that no competing interests exist.

sophisticated synthetic media by manipulating images, video, and audio content. This technology produces altered media of remarkable realism, often generating deceptive content indistinguishable from their authentic counterparts [2]. Indeed, deepfakes create a “simulation of the speaker in a hyper-realistic video” [3; p.16], representing people doing and saying things that have ever actually happened [4] by mimicking people’s facial expressions and voice modulations [5]. From a technical point of view, deepfake generation primarily relies on deep learning architectures, specifically Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs). The former were proposed by [6] employing probabilistic models to encode and decode data, offering a different approach to deepfake synthesis. The latter was introduced by [7] and comprised a generator and a discriminator network involved in an adversarial training process, resulting in highly realistic synthetic media. Deepfakes have become widely associated with applying deep learning techniques for face replacement [8], including facial reenactment and lip-syncing, all geared toward creating highly realistic videos. For instance, Deep Video Portraits [9] employ neural networks to transfer facial expressions from a prior source to another, producing lifelike visual outputs. Lip syncing, a critical component in audio-visual synchronization, is achieved through techniques like SyncNet [10] to enhance the inner coherence of the generated content. Recently, deepfake applications have gained significant popularity and have also been positively employed in various industries, including movies, gaming, social media, education, healthcare, material science, fashion, and e-commerce [11]. For instance, apps such as Deepware Scanner, Descript Overdub, RefaceApp, Avatarify, and Lil Miquela have been identified as noticeable for their distinctive features and varied functionalities. The proliferation of deepfake technology is receiving increasing attention in academia, including social sciences, humanities, computer science, political sciences, and law. Even though scholarly research is still lacking in integrating different perspectives on mechanisms involved in creating, consuming, and disseminating deepfakes [12]. The predominant matters of recent literature have been on deepfake negative implications and risks for media exposition, ethical concerns, and an overall necessity to increase public awareness regarding the potential misuse of AI [11]. There is a growing concern that users might soon be unable to discern content generated by machines from authentic ones, making them vulnerable to disinformation campaigns [13]. Experts worry about the malicious use of deepfakes in damaging societies [14]. Indeed, the use of deepfakes for spiteful purposes, such as nonconsensual pornography [15,16] (and political disinformation, undermines trust in institutions, media, online communication platforms, and societal values [17,18]). Some studies reported that exposure and sharing of deepfake videos could cause skepticism in considering new social media content [19–22] (Ahmed, 2021a; 2021b; 2023). In the political field, several factors heighten the threat of deepfake disinformation, such as human tendencies to be drawn to shocking content often present in deepfakes, contributing to broader audiences and facilitating dissemination [23]. Examples of deepfakes used for political propaganda were the fake video in which Obama swore at Donald Trump during an announcement in public [22] and the more recent fake video depicting Ukrainian President Volodymyr Zelenskyy issuing orders for his soldiers to lay down their weapons and cease fighting against Russia [24]. A study [8] highlighted university students’ main concerns regarding the spread of disinformation since malicious utilization of deepfake videos has been employed to create convincing fake news or manipulate public opinion. Other potential risks rely on privacy concerns since deepfake technology allows fake but realistic videos or audio recordings to be created without individual’s content, raising legal and ethical issues founding that privacy protection is the most sensitive factor [25,26]. Ethical concerns related to deepfakes include privacy protection, traceability, and informed question as factors that influence ethical acceptability directly and social acceptance indirectly. In contrast, perceived enjoyment directly affects social acceptance of

deepfakes by weakening the effect on ethical acceptability on social acceptance [26]. Also, the use of deepfakes for entertainment purposes could raise concerns about manipulating hyper-realistic digital representations of individuals' images and voices, which should be considered a fundamental moral right [27].

Scholars in cognitive psychology discussed individuals' mental processes underlying the fruition and detection of real or deepfake video. The mixture of realistic audio and visual cues could cause people to use their realism heuristic [28,29], in which visual cues precede signals from other senses in prompting responses and information storage [30,31] so creating even false memories [32,33]. Using heuristics facilitates cognitive efficiency since people select specific sensory inputs to process, disregarding other stimuli. Therefore, the level of attention paid affects the amount of information assimilated and processed [21] to prevent an overwhelming accumulation of information which could result in cognitive overload and, consequently, failure to comprehend and assimilate the content of the messages [34]. Individuals perceive audio and images as more closely mirroring the real world than text [22]. Within the metacognitive experience, fluency is central to comprehending why people believe false information. As suggested in [35], fluency implies that humans are more inclined to perceive messages as valid if they are familiar with them. This sense of familiarity causes a truthiness effect—i.e., a perceptual flow that facilitates easier assimilation of material, rendering it more believable [36]. Activated when images and audiovisual content prove more accessible and comprehensible than written texts, as delineated by [37], this metacognitive experience becomes instrumental in shaping people's cognitive task responses, notably in elaborating novel information. The technical realism of deepfake videos, mainly when depicting famous or widely recognized individuals, potentially compounds the preexisting concern that fluency can be achieved by invoking familiarity, regardless of the video's content accuracy. A third-person perception bias might also arise, and people are more inclined to believe that deepfakes affect others more than themselves and that they are better at distinguishing deepfakes than others. The third-person perception bias is less effective in children or adolescents due to their levels of cognitive development. However, it is more pronounced among adults with higher cognitive skills who are more skeptical of social media news and less inclined to evaluate deepfakes as accurate or share them. However, this bias does not necessarily align with people's ability to distinguish deepfake videos from real ones, as demonstrated by a deepfake detection test [21]. People outperform AI detection systems because of their ability to process faces holistically and visually [38]. Also, people may become excessively vigilant towards possibly manipulated media, overestimating the prevalence of deepfakes [39]. Overconfidence stands out as a prevalent and costly bias in people's decision-making. When people detect synthetic content, overconfidence could render people susceptible to manipulation. If individuals are confident in their ability to spot a deepfake but cannot, they may inadvertently engage with manipulated content. Moreover, analytic thinking and political interest are related in a positive way to correctly detecting deepfakes and are associated with the ability to discriminate fake news negatively [40]. People with high cognitive levels [21] interacting with people's cognitive biases on deepfake might exacerbate distrust in news and information disseminated by public figures [11,41]. In essence, trust constitutes an alternative decision-making mode, arising from willingly exposing oneself to vulnerability, acknowledging the inherent risk of betrayal, and giving precedence to someone's words over other forms of information [13]. Studies have underscored that media literacy education plays a crucial role in empowering individuals to critically evaluate the media they consume, including deepfakes. Interconnected with information and digital literacy, the media literacy aims to equip individuals with the skills to become informed and discerning consumers of media, enabling them to interpret messages and engage meaningfully with media content, even identifying biases [42].

Increased awareness and understanding of technology allow individuals to evaluate the media they consume, improving their ability to recognize and critically evaluate deepfake content [43,44].

Although many studies focused on the above-mentioned negative implications of deepfake, the current literature also reports the potential benefits and positive use cases of deepfake technology [17]. In a case study, deepfake technology was applied in educational settings to create realistic simulations or training scenarios for medical professionals who could practice and refine their skills in a virtual environment [45]. Besides, deepfake technology seems helpful in improving accessibility for individuals with disabilities since it can generate realistic sign language interpretation or assistive technologies for those with speech impairments, but researchers are also discussing potential risks [46]. According to [47], deepfakes could be used for creating simulated learning experiences, allowing students to practice skills in a safe and controlled environment, or for historical reenactments, bringing historical figures to life and allowing students to interact with them so gaining a deeper understanding of historical events. Besides, deepfakes can generate realistic conversations with native speakers, providing students with immersive language learning experiences [48]. For instance, the CereProc group recreated the authentic speech of John Fitzgerald Kennedy in July 1963 about the resolution to end the Cold War, based on previous speeches that he had given (<https://www.cereproc.com/en/jfkunsilenced>).

The present study

The current study aims to delve into the cognitive mental representation in people's perception of deepfake content, shedding light on the complex relationship between technology and psychological processes. Specifically, we applied thematic analysis using Tlab-10, a computer-assisted qualitative analysis software, to unveil patterns, correlations, and recurring themes of information discussed on Quora, an online forum by Quora Inc., founded in 2009, allowing users to publish questions and answers on specific topics. Questions and answers are grouped by topic, allowing users to vote or comment. Additionally, users can collaborate by modifying questions or answers provided by others. We opted for this platform due to its public nature and unrestricted access to media content. Like social platforms such as Twitter, it offers the advantage of not imposing character limits. Compared to Reddit, Quora has received less criticism for spreading misinformation [49].

The Theoretical and Conceptual Background section (Section 2). The Method section outlines the data collection and analysis procedures (Section 3). Results are subsequently discussed, highlighting prominent themes, and relating them to existing literature on deepfakes and human-computer interaction (Section 4). The Discussion reports the practical implications of the findings, emphasizing the significance of understanding public perceptions of deepfakes for comprehending their cognitive, social, and cultural implications (Section 5). Finally, Strengths and Limitations (Section 6) and Conclusion (Section 7) are reported.

Theoretical and conceptual background

Schema and media framing theory. The well-known schema theory [50–52] in brief affirms that, schemas are organized knowledge stored in the minds that people develop by interacting with the environment through the assimilation process [50,53]. They regard declarative - i.e., knowledge of what is an object, a fact, or an event - and procedural knowledge - i.e., knowledge of how to perform an action or behavior- [54]. Then, such cognitive schemas or knowledge units for a subject or event guide individuals to organize and process new information [55]. Every time people face new information acquisition, the perception associated is not

solely driven by external stimuli but is shaped by pre-existing knowledge structures [56]. Pre-existing knowledge could be adapted to new information through the accommodation process. Schemas also affect memory since they guide individuals' ability to pay attention to crucial information for encoding and help comprehend news by integrating it with prior knowledge [56,57]. In this way, the process allows the brain to manage its resources effectively, processing a few items but generating complex responses to them instead of processing a large amount of accessible information superficially [58]. Under schema theory [50–52], individuals could have pre-defined mental schemas or expectations about deepfake videos regarding information exchange. Therefore, schemas play a crucial role in attributing meaning to what happens around us and facilitating the evaluation, processing, and organization of a wide range of new information [55]. In [59], a pioneering exploration of framing is provided, so that media framing research is born. A media frame refers to a stable and socially shared system of categorization that influences how people perceive and behave in social contexts. A *frame* is a cognitive structure that shapes people's perception and mental representation of events. For instance, texts like newspaper clippings elucidate how primary frameworks function through keying, fabrication, and anchoring processes [59]. In the media framing theory, the active mental process of selecting frames and their outcomes is also emphasized [60]. According to a prominent perspective, framing entails selecting certain features of a perceived reality and highlighting them in a text "in a way as to promote a particular problem definition, causal interpretation, moral evaluation, and treatment recommendation" [60, p. 52]. In other words, framing influences the selection of events for news and how they are represented. By selecting and reporting news, the media leads people to focus on some problems and issues, not others [61]. People's reactions to framing in communication texts are primarily influenced by the standard schemas already present in their minds, which originate from various sources. For instance, a study underscored how exposure to framing strategies, particularly in the context of political campaign news, could evoke a recall of specific information strategies within individuals, consequently fostering the attribution of cynical motives to political actors [62,63]. This phenomenon embodies what is commonly referred to as a *priming effect*, wherein individuals' perceptions and evaluations are subtly influenced by the framing of information they encounter [64]. Media framing is central to ongoing research in communication. A *media frame* (i.e., graphical or visual, written or spoken) represents a tool that individuals use to provide context for a topic (i.e., event, issue, or people) which, throughout mediation, can be transmitted. Indeed, the concept of framing is a tool for shaping public discourse, particularly concerning the dissemination of information by media entities [63]. As stated, framing encapsulates a macro-attribute akin to an issue frame, wherein information sources strategically utilize various devices to meld and articulate opinions or preferences about a given situation. These frames may manifest in either a generic or topic-specific manner, thereby influencing the accessibility of information, either by emphasizing message salience or subtly biasing information processing mechanisms [63].

Examining media framing in the coverage of deepfake technology offers valuable insights into the evolving landscape of digital media regulation. By critically analyzing the framing strategies employed in news coverage, researchers can reveal the underlying narratives that shape public perceptions of deepfakes and inform discussions on regulatory interventions in the digital sphere. According to [65], media framing influences and reinforces collective views of reality. Public perceptions of situations or social groups shape the solutions deemed appropriate for addressing societal challenges, and in communication media frames are employed to align individuals with the contextual information included within the frame references, influencing their perceptions and behaviors regarding a specific topic [63]. In recent years, media framing and cognitive schemas have formed beliefs, expectations, and attitudes toward

technology [66]. A specific focus is addressed on people's cognitive ability to create their mental schemas to understand information spread through social media and the subsequent capability to distinguish real news from fake news. Through framing, news media can potentially increase people's awareness of a particular topic, drawing attention to actions and emphasizing potential solutions related to health and foreign policy [67,68]. Drawing upon expectancy-value theory [69] subsequent research articulated a comprehensive theory of opinion formation, elucidating the intricate interplay between journalistic framing and individual cognitive processes. This theoretical framework integrates the cognitive importance of information accessibility with two additional constructs, namely, availability and applicability [63]. According to this theoretical paradigm, the information presented within issue frames interacts dynamically with individuals' pre-existing knowledge and beliefs, eliciting accessible considerations that are readily recalled. Consequently, individuals consciously or subconsciously evaluate this information within their existing knowledge structures, shaping their overall attitudes and opinions on a given issue. Under the framework of cognitive psychology, authors examined how the valence-framing effects of information, whether in a positive or negative light, can systematically influence audience reactions. In [70] Levin and colleagues (1998) is proposed that the attribute-framing effect that occurs when positive framing elicits a more favorable response and negative framing leads to a less favorable response. This effect is thought to arise from the underneath psychological processes of information encoding and memory association. Specifically, the negative framing of an attribute can influence how information is encoded, potentially triggering an unfavorable memory association, while positive framing has the opposite effect. Additionally, it is suggested that the different encoding resulting from negative or positive valence frames may cause people to focus on the information differently. Moreover, the relationship between media frames and audience frames can be influenced by various factors, including social-cultural, organizational, individual, or ideological differences related to the issue [66]. Therefore, it is essential to investigate emerging valence frame trends in deepfake YouTube videos and their audiences. This will enable other researchers in the field of deepfake technology to advance their studies. Indeed, audience frames facilitate the processing of issue frames and play a constitutive role in forming opinions and expressions [71].

Thus, journalistic framing is a potent mechanism through which information is disseminated, shaping public discourse and perceptions on various socio-political issues.

In the era of deepfakes, information about the story's setting, characters, and timeline is twisted, painting a dystopian future picture, and envisioning a society controlled by altered content and disinformation. This usually happens when journalists narrate deepfakes as a politically and socially counterproductive phenomenon [71].

Method

Data collection

Data were extracted from Quora.com on November 20, 2023 (see [S1 File](#)). Specifically, we used the Quora internal search function to identify all questions related to deepfake by using "deep fake" or "deepfake" as search terms. The search identified 32 questions. We manually filtered these questions based on the following criteria: 1) Relevance: questions must explicitly mention "deep fake" or "deepfake" to ensure they are directly related to our research topic; 2) Content Availability: questions must have at least one answer to be included in the corpus.

This filtering process resulted in the exclusion of 4 questions that did not contain the search terms and 11 questions that did not have any answers, leaving us with 17 relevant questions.

We then used Octoparse software (Octoparse Data Inc) to extract the text and publication date of individual answers to each question. Octoparse is a free web scraping software that

takes unstructured data and text from websites and exports them to a structured data file. This focused approach ensured that the corpus consisted solely of questions directly relevant to the topic, thereby strengthening the meaningfulness and relevance of the data collected. By concentrating on questions that explicitly mentioned deepfakes, we aimed to capture answers covering a comprehensive range of perspectives and discussions surrounding this emerging technology. This methodological choice was essential in ensuring that the subsequent analysis accurately reflected public opinions and mental representations of deepfake. The final text corpus consisted of 166 answers to 17 questions. We included all answers to the selected questions, as each was directly relevant to the topic of deepfakes, ensuring that the corpus fully represented the perspectives and discussions related to the subject.

Answers were posted between February 10, 2018, and November 12, 2023. On average, questions received approximately ten answers (modal value = 5; min = 2; max = 24). The length of the answers was heterogeneous, ranging from three-word statements to a paragraph of 1,146 words (Mean = 145.24; SD = 174.60). This variability in answer length contributed to a rich dataset, capturing a wide range of perspectives on the topic of deepfakes. The study collected freely available public data from Internet forums. The data was accessed and analyzed following the platform's terms of use and all relevant institutional/national regulations. Data use complied with ethical guidelines for internet research [72]. The European Union General Data Protection Regulation 2016/679 allows for the use of anonymous data for research purposes under certain conditions. Since all analyses have been performed on public and anonymized data, no institutional review board approval was required for the use of this database or the completion of this study.

Data pre-processing

Before analyzing data, the text corpus underwent accurate pre-processing procedures to ensure consistency and accuracy across the dataset. Specifically, manual intervention was undertaken to rectify typographical errors and standardize terminology usage. Notably, terms such as "deep fake" or "deepfake" or "deep fakes" or "deepfakes" were consistently recoded as "deep_fake" to foster uniformity throughout the corpus. Furthermore, terminological standardization was performed to avoid ambiguity and implement more robust analytical processes. Noteworthy efforts included the substitution of abbreviated terms such as "AI" with "artificial_intelligence" and "ML" with "machine_learning," ensuring clarity and precision in subsequent semantic analyses. Naming conventions were applied to ensure consistency in the representation of people within the corpus. Notably, proper nouns and surnames, such as "Queen Elizabeth" or "George Lucas," were formatted uniformly as "queen_elizabeth" and "george_lucas," respectively, facilitating coherent identification and analysis. The corpus was primed for subsequent analysis with enhanced uniformity and accuracy by systematically recording terms and addressing typographical inconsistencies. Finally, the answers' publication dates were categorized as "Older answers" (i.e., more than three years old) or "Newer answers" (i.e., two years old or less).

Analysis method

The individual answers were analyzed through T-Lab 10 [73], a computer-assisted qualitative analysis software that included linguistic and statistical tools for text mining and thematic analysis. Upon importing the corpus into T-Lab, each answer underwent an automated segmentation process into elementary contexts, defined as short paragraphs of similar length. This resulted in a total of 566 elementary contexts, which provided a manageable unit for analysis. The coding process involved the identification of significant words and lemmas within

the corpus. The coding was performed through automated lemmatization, which reduced words to their base forms, allowing for a more straightforward analysis of frequency and co-occurrence. Lemmatization is the process of reducing words to their base or dictionary form (the lemma), based on their intended meaning in context. For example, the words “creating”, “created”, and “creates” would all be reduced to the lemma “create.” This allowed us to group together inflected forms of a word and analyze them as a single item. Following our main aim, the analytical process consisted of three main phases:

1. Identification of the most frequent lemmas used in deepfake-related Quora answers. The primary objective of this analysis is twofold: first, to discern the predominant lemmatized forms of words utilized within the deepfake-related Quora corpus, and second, to quantify their respective frequencies of occurrence. Specifically, after lemmatization, we conducted a frequency analysis to determine the most frequently occurring lemmas in the corpus. This analysis was performed automatically by T-Lab, which counted the number of times each lemma appeared across the 566 elementary contexts that made up the corpus. To identify the most used lemmas, we focused on the frequency of each lemma in the corpus. We ranked all lemmas based solely on their occurrence counts, selecting the top 20 lemmas that were most frequently used in discussions related to deepfakes.
2. Co-occurrence analysis through “word association”. Through this analysis, we aimed to identify words frequently appearing with the lemma “deep_fake” within elementary contexts and identify semantic relationships and associations between terms. We utilized T-Lab’s automated tools to calculate co-occurrence frequencies, focusing specifically on the instances where the lemma “deep_fake” co-occurred with other lemmas. The analysis generated tabular and graphical representations that highlighted these co-occurrences, allowing us to visualize the relationships between terms. To quantify the strength of these associations, we employed the Chi-square test as a measure of association. This statistical test helped us determine whether the observed co-occurrences of “deep_fake” with other lemmas were significantly higher than what would be expected by chance. By focusing solely on the co-occurrences of “deep_fake” with other lemmas, we aimed to identify key word pairs that reflect the most salient themes and sentiments in discussions about deepfake technology. Specifically, by utilizing word associations, we construct tabular and graphical representations highlighting the co-occurrence of the lemma “deep_fake” with other words, offering insights into the complex network of relationships it has within the corpus and providing an in-depth understanding of how these terms interrelate in the discussions surrounding deepfake technology.
3. Identification of coherent semantic clusters defined by distinctive word patterns, commonly called themes. This involved conducting Singular Value Decomposition (SVD) and successively hierarchical cluster analysis (i.e., Principal Direction Divisive Partitioning and K-means method) [74,75]. SVD serves as a tool for reducing dimensionality, revealing latent dimensions that underlie semantic similarities among words. By applying SVD, we were able to identify patterns in the data that might not be immediately apparent, thus facilitating a more nuanced understanding of the relationships between terms. Subsequently, cluster analysis utilized the outcomes of SVD to pinpoint semantic clusters, or themes, characterized by specific word arrangements. In this step, we assessed the coherence and distinctiveness of the clusters, ensuring that each theme accurately represented a significant aspect of the data. This methodological approach adheres to a ‘bottom-up’ methodology rooted in induction, wherein the themes extracted are closely tied to the empirical data. This inductive approach allowed us to remain grounded in the actual responses from participants,

ensuring that the themes were reflective of the public discourse surrounding deepfake technology. By closely linking the themes to the data, we aimed to capture the complexities and nuances of public perceptions.

- Testing the associations of thematic clusters with answers' publication data. We analyzed the relationship between thematic clusters and answer publication dates using a Chi-square test and a residual analysis. The residual analysis helped identify significant associations between themes and publication dates by considering: a) standardized residuals greater than 1.96, indicating that the number of observed values was significantly greater than expected, and b) residuals below -1.96 , indicating that the number of observed values was significantly less than expected.

Results

Key Lemmas analysis: Unveiling trends in deepfake discourse

In examining the deepfake corpus, we focused on analyzing the 20 most prevalent critical lemmas aimed at discerning predominant trends within the discourse. As shown in Table 1, the frequencies attributed to each lemma provide significant insights into the main aspects that users primarily focus on while discussing deepfake technology on Quora.

Unsurprisingly, “deepfake” emerges as the predominant lemma, since it was one of the main keywords we applied on researching inside the Quora community. The high prevalence of “video” underlines the significant emphasis on the visual aspect of deepfake content and its implications within the discussions. At the same time, the frequency of “fake” highlights possible concerns and the general awareness surrounding the potential deceptive nature of deepfake-generated content. Other most frequent lemmas, such as “create” and “technology,” underline the active generation and production of deepfake content, respectively, reflecting an interest in the creative dimensions of this technology and the users' interest in understanding the underlying technical aspects and advancements associated with deepfake.

Co-occurrence analysis: Unveiling semantic relationships

In the specific context of our analysis, co-occurrences of words were computed within elementary contexts defined during the corpus importation phase. This robust approach ensures that the evaluation is conducted at granular levels, allowing us to capture the nuanced associations

Table 1. The most frequently used lemma.

Lemma	Frequency	Lemma	Frequency
deep_fake	515	person	80
video	293	look	79
fake	146	content	70
create	144	audio	60
technology	135	news	56
image	110	good	55
people	98	voice	47
face	92	believe	44
artificial_intelligence	91	app	43
use_to	82	help	43

<https://doi.org/10.1371/journal.pone.0313605.t001>

that shape the semantic landscape of the key lemmas in our deepfake corpus. Specifically, we assessed associations of the lemma “deep_fake” with other lemmas (Table 2).

The identified word pairs offer significant insights into public perceptions of deepfakes. For instance, the strong association between “deep_fake” and “video” suggests that discussions predominantly focus on the visual implications of this technology. Similarly, the co-occurrence with “create” highlights the perception of deepfakes as tools for content generation, which can be both innovative and concerning. Terms like “artificial_intelligence” and “technology” further emphasize the advanced technological underpinnings of deepfakes, indicating that users are aware of the sophisticated methods involved in their creation. Conversely, the presence of words such as “fake,” “news,” and “manipulation” reflects public anxiety regarding the potential for misinformation and ethical dilemmas associated with deepfake technology. This duality in word associations illustrates the complex nature of public sentiment, where excitement about technological possibilities coexists with significant concerns about trust and authenticity in media.

In details, as also shown in Fig 1, users associated with “deep_fake” lemmas related to the creation (i.e., “create”) of synthetic media, mainly “video” or “image” or “audio,” through “artificial_intelligence” and new “technology.” For example, users wrote: “deep_fake - a combination of deep_learning and fake is a term for videos and presentations enhanced by artificial_intelligence to present falsified results. One of the best examples of deep_fake involves videos of celebrities, politicians or others saying or doing things that they never actually said or did.” or “deep_fake can create numerous possibilities and opportunities for all, regardless of who they are and how they interact with the world around them. What is deep_fake technology? Simply put, deep_fake is a technology that easily lets you make and create the realistic-looking digital avatar of any real person.”

Thematic analysis: Unveiling the discursive threads in the quora corpus on deepfake

As a final analysis output, we identified eight main themes within the deepfake corpus, presented in Table 3. This table displayed the list of lemmas characterizing each cluster. Each theme summarizes a unique facet of the discourse, providing insights into different perspectives, concerns, and narratives in the discussions on Quora. The following sections describe

Table 2. Word association analysis for “deep_fake” within the Quora corpus.

Lemmas	Coeff.	EC(B)	EC(AB)	χ^2	P
video	.552	207	140	21.216	< .001
create	.470	110	87	32.151	< .001
technology	.470	110	87	32.151	< .001
artificial_intelligence	.415	79	65	27.706	< .001
use_to	.367	67	53	17.915	< .001
content	.343	58	46	15.495	< .001
fake	.342	99	60	1.552	.213
image	.338	85	55	3.848	.050
audio	.315	52	40	11.171	.001
person	.309	65	44	4.819	.028

Note: Coeff. = value of the coefficient; EC (B) = total number of elementary contexts including the associated lemma (B); EC (AB) = total number of elementary contexts in which lemmas A and B appear together (i.e., co-occurrences). This table displayed only the first ten lemmas B with the higher coefficients.

<https://doi.org/10.1371/journal.pone.0313605.t002>

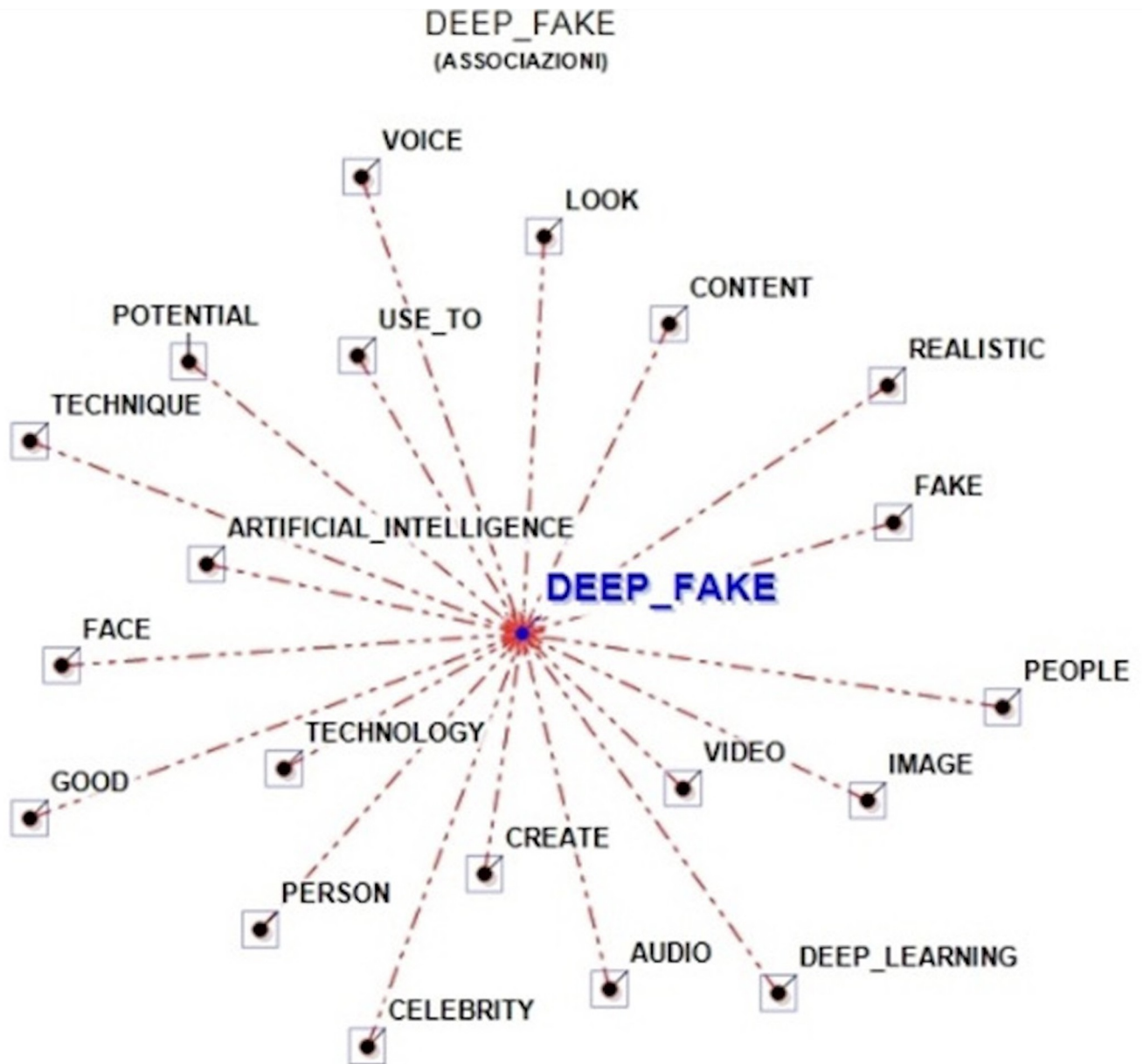


Fig 1. Graphical representation of the word association analysis.

<https://doi.org/10.1371/journal.pone.0313605.g001>

these themes, revealing the rich layers of meaning embedded within the deepfake discourse on Quora.

Theme 1: Deep Fake Generation Techniques

This theme focuses on the techniques and technologies of creating deep fakes, particularly pinpointing deep learning and generative adversarial networks (GANs). The lemmas highlight crucial elements such as deep_learning, fake, generator, discriminator, technique, clip, and base. The context emphasizes the process of deep fake creation, where a generator produces fake video clips, and a discriminator distinguishes between real and fake content. The theme

Table 3. The eight identified themes.

	Theme 1: Deep fake generation technique	Theme 2: Detecting Deep fake anomalies	Theme 3: Deep fake creation apps	Theme 4: Ethical reflection and responsible use
total occurrence (%)	71 (15.88)	54 (12.08)	23 (5.15)	65 (14.54)
lemmas (occurrence)	deep_learning (29)	able_to (21)	app (19)	potential (28)
	fake (62)	light (13)	google (6)	software (17)
	generator (17)	sign (17)	swap (9)	danger (14)
	discriminator (14)	shadow (11)	easy (8)	misinformation (14)
	technique (19)	blurry (9)	offer (5)	important (14)
	clip (16)	background (11)	play (5)	technology (41)
	base (12)	edge (8)	China (3)	use_to (30)
	portmanteau (7)	skin (7)	facial (11)	purpose (14)
	create (48)	tooth (7)	feature (8)	spread (15)
	hoax (6)	original (9)	photo (5)	responsibly (5)
	network (7)	perfectly (9)	expression (7)	create (41)
	refer (8)	noise (6)	look (14)	privacy (7)
	synthesis (8)	strange (7)	result (5)	risk (7)
	image (36)	inconsistency (9)	great (4)	associate (4)
	alter (11)	absence (6)	allow (5)	hashtags (4)
	term (12)	glitch (4)	spot (5)	job (4)
	synthetic_media (8)	inconsistent (5)	creation (2)	malicious (8)
	generative_adversarial_network (4)	telltale (5)	color (2)	manipulate (10)
	revenge (6)	double (4)	man (2)	increasingly (6)
	celebrity (16)	artifact (4)	moment (2)	fingerprint (4)
	Theme 5: Entertainment and technological innovation	Theme 6: Threats to information integrity and societal impact	Theme 7: Facial mapping and superimposition	Theme 8: Facial cues and unnatural movements
total occurrence (%)	79 (17.67)	63 (14.09)	18 (4.03)	74 (16.55)
lemmas (occurrence)	actor (20)	information (12)	target (89)	eye (19)
	film (20)	election (10)	model (13)	emotion (9)
	movie (17)	false (16)	lip (7)	lack (7)
	entertainment (14)	reputation (6)	train (59)	talk (10)
	tid (9)	believe (17)	closely (3)	facial (20)
	star (8)	impact (6)	feature (8)	point (9)
	think (17)	political (9)	syncing (3)	natural (8)
	social (9)	threat (7)	zoom (4)	movement (14)
	medium (16)	democracy (4)	person (12)	pay (8)
	famous (7)	candidate (4)	common (3)	awkward (5)
	industry (8)	serious (5)	superimpose (3)	somebody (5)
	ban (5)	people (24)	capture (2)	attention (8)
	replace (12)	harm (6)	practice (2)	replicate (4)
	mimicry (4)	world (7)	screen (2)	position (6)
	queen_elizabeth (4)	reality (5)	slow_down (2)	facts (5)
	community (6)	fool (6)	training (3)	skeptical (5)
	tv (6)	damage (4)	facial (7)	sound (5)
	viewer (6)	mass (4)	expression (5)	blink (7)
	artificial_intelligence-generated (5)	truth (6)	latent (2)	focus (5)
	message (5)	cause (5)	data (4)	body (10)

<https://doi.org/10.1371/journal.pone.0313605.t003>

also underlines the possible implications of deep fake technology, including its misuse in generating fake celebrity content, revenge porn, and fake news. Overall, the theme provides insights into the mechanics of deep fake generation and its impact on media synthesis and manipulation. An example evidencing the public discussion about technical GAN technologies related to deepfake is reported in the following: “Basically, the **generator creates a fake video clip** and then asks the **discriminator** to determine whether the **clip** is real or **fake**. Each time the **discriminator** accurately identifies a video **clip** as **fake**, it gives the **generator** a clue about what not to do when **creating** the next **clip**. Together, the **generator** and **discriminator** form something called a **generative_adversarial_network**”.

Theme 2: Detecting Deep Fake Anomalies.

This theme is about identifying anomalies and irregularities in deep fake content. The lemmas focus on characteristics such as *able_to*, *light*, *sign*, *shadow*, *blurry*, *background*, *edge*, *skin*, *tooth*, *original*, *perfectly*, *noise*, *strange*, *inconsistency*, *absence*, *glitch*, *inconsistent*, *tell-tale*, *double*, and *artifact*. The context emphasizes various signs and indicators that may reveal the presence of a deep fake. These signs include strange lighting or shadows, blurry or distorted features, skin tone or teeth abnormalities, inconsistent noise or audio, and background inconsistencies. The theme highlights the importance of recognizing anomalies such as artifacts, glitches, or double edges in the visuals and the absence of blinking or unnatural facial expressions. It also addresses the role of critical thinking in evaluating videos and suggests reporting deep fakes to platforms with policies against them. Overall, the theme provides insights into the visual and auditory cues that may indicate content manipulation through deep fake techniques, as reported in the example: “Strange **lighting** or **shadows**: deep_fake can sometimes have **strange lighting** or **shadows**, as the artificial intelligence model may not be able to recreate the **lighting** conditions of the **original** video perfectly. **Artifacts** or **glitches**: deep_fake can sometimes have **artifacts** or **glitches**, as the intelligence model may not be **able to perfectly** blend the two images”.

Theme 3: Deep Fake Creation Apps.

In the following example “Reface: ✓✓✓ Reface: Face **Swap** artificial_intelligence **Photo App**— **Apps** on **Google** Play Create & share fun face-swapping videos in seconds with amazing artificial_intelligence technology. // **play. Google. com/store/apps/details? id = video. reface. app&hl = en&gl = US** Reface is another **great** deep_fake **app** that **offers** an impressive **feature** set” we note how this theme focuses on applications used for creating deep fake content and the associated concerns, with lemmas like the *app*, *google*, *swap*, *easy*, *offer*, *play*, *China*, *facial*, *feature*, *photo*, *expression*, *look*, *result*, *great*, *allow* *spot*, *creation*, *color*, *man*, and *moment*. The primary context describes different apps on platforms like Google Play and Apple, offering features like face swapping, artificial intelligence-driven facial expression analysis, and creating realistic deep fake content from images and videos.

Theme 4: Ethical Reflection and Responsible Use.

This theme underlines the possible ethical dimensions and responsible application of deepfake technology. The key lemmas include “potential,” “software,” “danger,” “misinformation,” and “technology.” Discussions focus on the dual nature of deep fake technology, recognizing its potential for both positive and negative purposes. Concerns are raised about using deep fakes to spread misinformation, and users are advised to be aware of the associated risks. Tips for responsible use are provided, emphasizing transparency and protecting privacy. The sophistication of deep fake software and its evolving nature are underlined by stressing the relevance of increasing public awareness about deepfake. Overall, the theme emphasizes the need

for a balanced understanding of the risks and benefits associated with deep fake technology, as we can see in the example: “This raises concerns about the **potential** for deep_fake to be used to **spread misinformation** or to **manipulate** people. It is **essential** to be aware of the **potential dangers** of deep_fake **software** and to use it **responsibly**. If you are considering using deep_fake **software**, it is **crucial** to be mindful of the ethical implications of this technology”.

Theme 5: Entertainment and Technological Innovation

This theme underlines the impact of deep fake technology on the entertainment industry and society. Key lemmas include “actor,” “film,” “movie,” “entertainment,” and “TID.” Discussions highlight the transformative possibilities of deep fake in the movie industry, envisioning scenarios where any actor can be seamlessly replaced or chosen post-production. Examples are provided, such as the recreation of iconic movie scenes with different actors. TID (The Indian deep_faker) is introduced as a prominent player in creating artificial intelligence-generated content for social media. The ethical implications of deep fake in entertainment are acknowledged, with perspectives on its positive and negative aspects. Overall, the theme emphasizes the evolving role of deep fake in shaping the entertainment landscape and its broader societal implications as we can see in the example: ✓ Not only can any **actor** be **replaced** in any past **film**, but **films** can be planned and produced in the future to replace the **actor**. ✓ Imagine a new **movie** experience, in which the **film** was shot and edited with any unknown **actor**, but you, the **viewer**, can **choose** later who you want to **star** in the **movie**.

Theme 6: Threats to Information Integrity and Societal Impact.

This theme revolves around the potential threats of deep fake technology to information integrity, elections, and societal stability. Key lemmas include “information,” “election,” “false,” “reputation,” and “believe.” Discussions emphasize the profound impact deep fake videos could have on elections, public opinion, and the economy. Concerns about the harm to reputations and the use of deep fake news to spread fake news, especially in the context of political manipulation, are highlighted. The narrative touches on the broader societal implications, discussing the erosion of trust in information sources and the challenges of differentiating between reality and manipulated content. The theme also acknowledges the dual nature of deep fake technology, which can be used for both positive and nefarious purposes, with a call for vigilance and awareness. This emerges indeed in the example: “I **believe** most of them present whatever **information** that supports their **political** narrative. There is plenty of MSMEDIA parroting the same **information** repeatedly. Much of it was **false**. Advertising has proven that the **masses** are affected positively by advertising. Most **people** I **believe** the half truths of advertisements”.

Theme 7: Facial Mapping and Superimposition. This theme revolves around facial mapping and superimposition facilitated by deepfake technology. Key lemmas include “target,” “model,” “lip,” “train,” “feature,” and “syncing.” Discussions highlight the feature extraction process from a targeted individual’s face, emphasizing the importance of capturing facial expressions, head movements, and speech patterns. Terms like “feature,” “syncing,” and “superimpose” highlight the focus on replicating facial expressions and movements. The elementary contexts describe how AI models, once adequately trained, can map the features of a target person onto different videos or images, achieving a realistic superimposition of faces onto various scenarios. The discussions emphasize the importance of feature extraction and the challenges related to unnatural lip-syncing and facial expressions in deepfake videos. The theme also discusses practical aspects, including the need for data to train the model and standard practices in deep fake creation. There is also a nod to the advancements in detection techniques and the importance of vigilance in identifying signs of deep fakes. This is for instance

shown, in the example: “Feature **Extraction**: The artificial intelligence model learns to extract key **features** from the **target** person’s face, such as facial expressions, head movements, and speech patterns. These features are crucial for ensuring that the generated content **closely** mimics the target’s behavior”.

Theme 8: Facial Cues and Unnatural Movements.

The eighth theme delves into the intricate details of facial expressions and movements in the context of deepfake technology. Key lemmas such as “eye,” “emotion,” and “lack” underscore the focus on replicating natural facial cues. The elementary contexts highlight the challenges in replicating authentic eye movements and facial expressions and the significance of paying attention to specific facial features such as cheeks, forehead, and eyebrows. Discussions emphasize the difficulty in mimicking natural blinking and eye movements, pointing out these challenges as potential red flags for detecting deepfake content. Common signs of deep fake videos are explored, including unnatural facial expressions, awkward facial positioning, and inconsistent body movement. The theme also addresses the challenges of replicating specific actions, such as blinking, in a natural way. Techniques for spotting facial morphing or image stitches are discussed, focusing on identifying emotions that may not align with the spoken content. The importance of scrutinizing facial feature positioning and looking for signs of a lack of emotion in videos is emphasized.

Overall, as shown in the example: “**Pay attention** to the **face**. . . . **Pay attention** to the cheeks and forehead. . . . **Pay attention** to the **eyes** and eyebrows. . . . **Pay attention** to the glasses. . . . **Pay attention** to the **facial** hair or **lack** thereof. . . . **Pay attention** to **facial** moles” this theme guides individuals to detect potential deep fake videos by examining specific visual and emotional cues.

Associations between cluster membership and publication date. The result of the Chi-square test highlighted significant associations between themes and publication date ($\chi^2(7) = 34.01, p < .001$). Specifically, as displayed in Fig 2, the first theme is more likely in older answers (SR = 3.92). In contrast, the second and the fourth themes are more likely in more recent answers—i.e., SR = 2.07 and SR = 3.89, respectively.

Discussion

The current study investigated people’s mental representation of deepfake and its related contents in the human-computer interaction framework, applying a thematic analysis to public discussion on Quora online forums. The qualitative examination of discussions through Quora provided insights into how individuals perceive and engage with deepfake content in an online forum setting. This approach facilitated exploring the diverse themes of responses to deepfakes, revealing users’ expectations, concerns, and attitudes toward synthetic media.

Results evidenced eight main themes that differ in their specific contents and a temporal frame effect, showing significant differences between older and recent discussions around each theme. Public discussions on Quora focusing on Theme 1 - Deep Fake Generation Techniques decreased from 25% to 11% from past to present, indicating that people have become more aware about the technical development of deepfake technology. Nowadays, creating artificial visual media has become remarkably easy. The rising popularity of deepfakes is due to their convincingly realistic videos and user-friendly interface, which are accessible to individuals with diverse levels of computer proficiency. Generating convincing fake videos is becoming progressively simpler due to technical advancements in AI. Now, all it takes is a target person’s photo or a brief video to produce remarkably realistic altered content [76]. Similarly, discussions on Theme 2 -Detecting Deep Fake Anomalies, which collects people’s opinions about

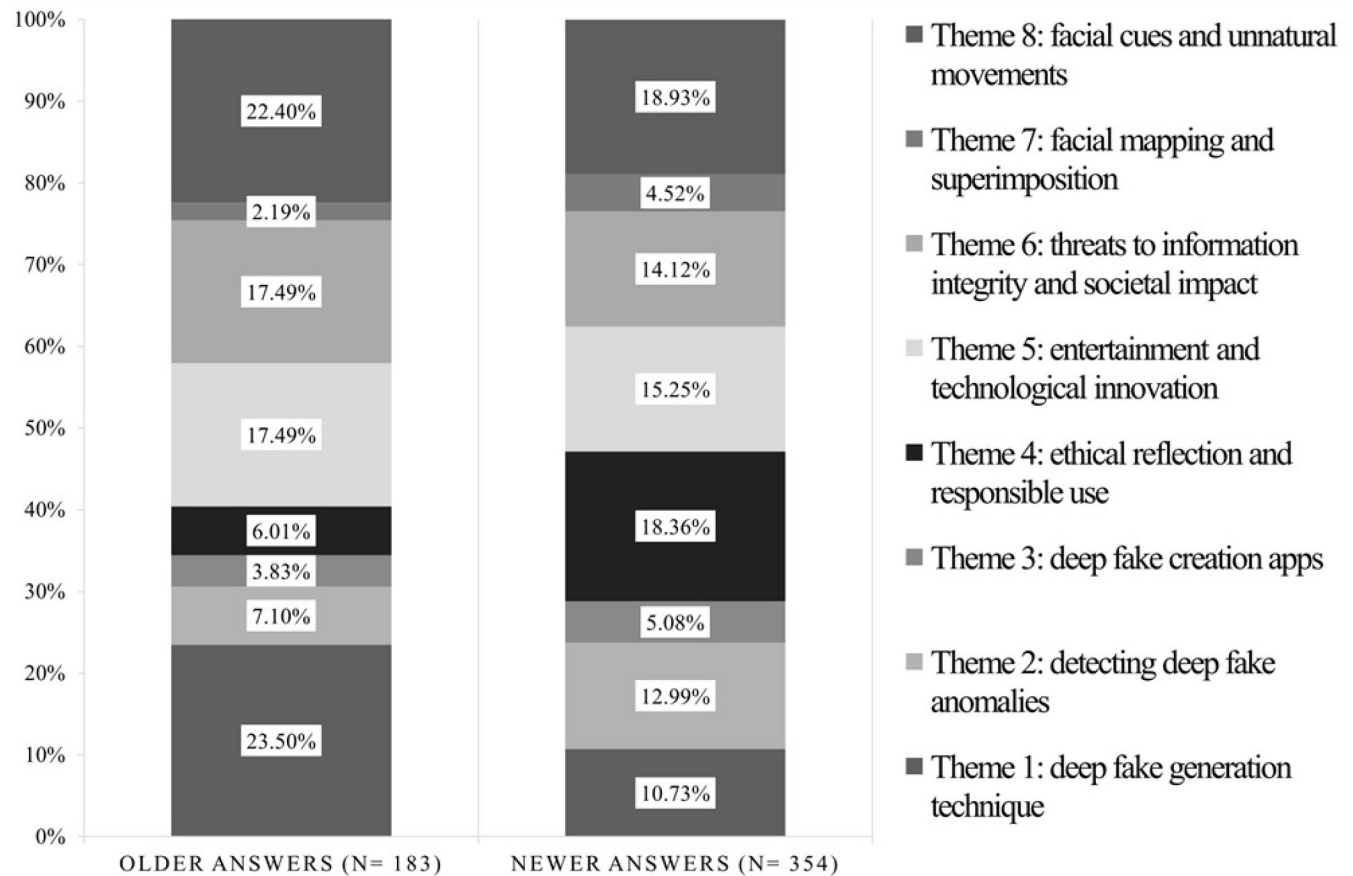


Fig 2. Percentages of elementary contexts classified into each cluster for older and newer answers.

<https://doi.org/10.1371/journal.pone.0313605.g002>

media inconsistencies and irregularities in deepfake data, increased from 10% to 14% from older to recent times. The primary emphasis in people's discussions is on strategies and methods for deepfake detection, which is focused on identifying anomalies to detect artifacts and traces originating from the underlying AI generative process [77]. In deepfake detection, visual cues remain valuable, orienting people's ability to detect real vs. fake media, but determining the authenticity of a video goes beyond just visual processing. From a psychological perspective, we could argue that deepfake detection involves considering both the technological context and critical thinking skills, so adapting beliefs based on new information [38,78].

The prevalence of discussions related to Theme 3 *Deep Fake Creation Apps* varied from 3% in old posts to 6% in recent ones, suggesting a growing interest in applications associated with deepfake technology. Characterized by the ability to manipulate and generate realistic audiovisual content, deepfake applications inherently possess a novelty factor. Recent studies on media technologies suggest that individuals could be initially drawn to the uniqueness and creativity associated with deepfakes, perceiving them as a novel form of entertainment [79]. In this sense, deepfakes could enhance the well-known *novelty effect* in people, defined in psychology as a phenomenon that describes the initial surge of interest and attention individuals show when exposed to something new or novel. Mainstream psychology used it in the context of research related to human perception [80] elving into the role of curiosity and arousal and exploring how novelty can contribute to increased interest and attention. Besides, the novelty effect was related to learning and behavior. Hull's work on habit-family hierarchy and maze

learning touches upon the role of novel stimuli in influencing learning and behavior [81]. The novelty effect has also been associated with dispositional traits such as curiosity [82] and sensation-seeking [83]. Regarding media psychology, the novelty effect has been observed in various technological advancements, influencing the perception and adoption of novel forms of media and entertainment [29,84], affecting whether they are seen as a source of entertainment or perceived as a potential threat.

Regarding Theme 4: *Ethical Reflection and Responsible Use*, results show that discussions significantly increased from 5% in older discussions to 19% in recent ones, indicating a heightened awareness and discussion of ethical implications surrounding deepfake technology. Deepfakes are exploited as tools for unethical practices in fields such as pornography [85,86]. Moreover, scholars highlighted the dangers and harms related to deepfakes [11]. Although deepfakes have received recently academic attention it is increasingly evident that they offer benefits and opportunities in social and medical fields (i.e., assisting individuals with Alzheimer's disease in interacting with a younger face they may recall). In 2020, a deepfake video revived a victim of the Parkland school shooting in 2018, advocating for gun safety legislation [87] (Diaz, 2020). This highlighted the potential of deepfakes for promoting pro-social causes. Evidence of positive Deepfake technology use is scarce, but it has been demonstrated to have the potential for positive employment. Implementing deepfakes in *FakeForward*, which refers to models involving peers, has demonstrated that desirable behaviors and skills are encouraged, which increases performance and confidence, among others. It is assumed that when selecting video material to use with *FakeForward*, it is essential to choose content that showcases activities contributing to an individual's positive development while ensuring protection from damage [88]. The current psychological literature evidences the framing effect, which relies on a strict interdependency between how technology is portrayed in the media and people's mental representations and beliefs about it [89]. The framing effect could influence cognitive processing, emphasizing the positive aspects associated with the novelty of deepfakes. As discussed, deepfakes can potentially foster trust and prepare individuals for the digital era. It could bolster collective critical thinking, mitigate susceptibility to misinformation, and encourage rigorous source verification. This, in turn, facilitates a purposeful transition from instrumental rationality to a socially informed trust paradigm in the digital age [13].

Results about Theme 5: *Entertainment and Technological Innovation* evidenced that people's interest in using deepfake for entertainment decreased slightly from 19% to 15%, suggesting a slight decrease in interest in the specific topic. Numerous Deepfake applications exhibit creativity, educational value, and amusement, yet they are often overlooked in the literature, focusing on the negative aspects [90]. Deepfakes have been suggested to personalize and make films, video games, and other media by incorporating one's face onto characters. For instance, in the trailer for the film *Gemini Man* in 2019 starring Will Smith, deepfake technology was employed in an old clip of Will Smith from the television series *The Fresh Prince of Bel-Air*, in which he talked about the movie. They also can potentially supplant CGI (Computer-Generated Imagery) in the film industry [88]. This technology is often controversially employed to bring actors back to life. As in this case, research suggests that media framing is crucial in shaping individuals' cognitive schemas regarding technology [66]. Different frames, such as innovation or risk frames, can influence how people acts. Iyengar (1991) In [91], it is discussed how media frames can activate existing cognitive schemas, influencing individuals' interpretation of information, and, in turn, this interaction contributes to the formation of people's beliefs and attitudes. Research examined the framing of technology in news media, arguing that positive framing tends to enhance beliefs in the benefits of technology, while negative framing emphasizes risks and potential drawbacks [92]. Recent examination of media framing in the context of AI underscores the role of media framing in shaping cognitive schemas related to

the societal implications of AI [93]. Prior literature also focused on how media significantly influences public perception and adoption of new technologies. Positive portrayals in the media can create an optimistic view of technological advancements, fostering enthusiasm and willingness to adopt innovations. In contrast, negative portrayals in the media could negatively influence public mood and perceptions [45] (Chen, 2004). Then, we could also hypothesize that deepfakes could be more likely interpreted by people as a novel and creative form of content rather than a potential threat when presented in an entertainment framing.

Moreover, the results of the current thematic analysis also evidenced *Theme 6 - Threats to Information Integrity and Societal Impact*, which collects discussions about *potential* threats of deepfake, evidencing an increasing trend from 9% in older discussions to 14% in recent ones, so indicating a growing concern people have regarding the potential misuse or threats posed by deepfake technology. As advanced tools for crafting misleading narratives, Deepfakes pose a significant risk of spreading false information. This perpetuation of falsehoods can prompt individuals to propagate rumors unknowingly or even intentionally, contributing to the spread of misinformation. The influence of misinformation on melding political perspectives, as the emergence of deepfakes carries significant implications for shaping political convictions and amplifying societal divisions [94]. Misinformation manifests in various forms, from isolated audio clips to fake news and low-quality manipulated media, such as cheap fakes [19]. Anyway, the deleterious impact of social media on democratic processes surpasses conventional misinformation [11]. Take, for instance, the emerging prevalence of deepfakes, which pose a significant and hostile threat, potentially leading to an "information apocalypse." In such a scenario, distinguishing between fact and fiction becomes increasingly challenging for citizens [11]. However, when individuals become aware of the falsehoods or face social repercussions for spreading them, they may seek to distance themselves from the rumors or halt their involvement in their propagation [95]. Deepfakes can be a potent tool for creating false narratives, making it imperative to study how exposure to such content influences the formation of false beliefs. From this point of view, deepfakes have the potential to shape political beliefs and exacerbate social polarization [95]. The role of misinformation in influencing political attitudes, understanding how deepfakes contribute to these dynamics is essential for comprehending the broader societal impact on political discourse and social cohesion].

Also, results on *Theme 7: Facial Mapping and Superimposition* evidenced an enhancement in discussions about facial mapping from 1% to 3% in more recent discussions, suggesting an imperceptible growing interest in the technical aspects of facial manipulation. However, it is noteworthy that discussions about *Theme 8: Facial Cues and Unnatural Movements* have decreased from 19% to 12% during the same period. This appears to be a shift towards a greater interest in the technical aspects of facial manipulation rather than the analysis of facial expressions themselves as well as tools for face swapping improved over the time. In this context, the challenges lie in accurately reproducing genuine eye movements and facial expressions. It underscores the necessity of paying attention to specific facial features such as cheeks, forehead, and eyebrows, as well as the challenge of mimicking natural eye movements, citing these difficulties as potential red flags for identifying deepfake content. Moreover, as emerged from recent studies [96], understanding users' eye movements and reactions to deepfake content is crucial for enhancing the realism and human-likeness of deepfakes, which can impact user trust and engagement. Indeed, a key precursor to lead people to accept the potential benefits of deepfake is overcoming the so-called uncanny valley effect [97]. This effect usually brings people to experience a feeling of discomfort when humanoid robots are highly realistic [96].

Strengths and limitations

Overall, this study has several strengths. First, it represents the first attempt to systematically identify and figure out themes related to deepfake content perception. Our qualitative approach allowed for an in-depth exploration of cognitive mental representations involved in people's perception of deepfake content. Specifically, this study contributes to the literature by comprehensively exploring the discourse surrounding deepfake technology. This work provides insights into the most frequent lemmas, semantic relationships, and themes related to this relevant topic through co-occurrence analysis and semantic clustering. This approach allowed for a better understanding of the diverse perspectives and narratives surrounding deepfake technology. Furthermore, by testing the associations of thematic clusters with answer publication dates, we gained insights into the temporal dynamics of deepfake-related discourse, highlighting how cognitive representations of deepfake might have evolved.

However, it is essential to acknowledge some limitations inherent in this study. Firstly, our approach is solely qualitative and does not incorporate quantitative data, limiting the depth of our analysis and the generalizability of our findings. Secondly, one potential limitation is the size of the text corpus analyzed. The final corpus consisted of 166 answers to 17 questions. While this may seem small, it is in line with recommendations for thematic analysis, as 100–200 items is a sufficient sample size for a small-scale thematic analysis project [98]. Given the exploratory nature of this study and the emerging nature of the deepfake phenomenon, the sample size is appropriate for the research objectives. It is important to note that it was not possible to increase the sample size beyond the 166 answers, as this represented the total number of relevant questions and answers available on the Quora platform at the time of data collection. Future research could expand on these findings by analyzing a larger corpus of data from different sources. Additionally, using Quora data might lead to potential biases in online platforms, such as the lack of demographic information about the users contributing to the discussions. Furthermore, we should also consider that users who contribute to discussions on Quora about deepfake technology may not represent a random population sample. Thus, while our study offers valuable insights into the cognitive processes underlying deepfake perception, the generalizability of its findings may be limited to Quora's users and may not fully capture the variety of perspectives on this topic.

In summary, while this study offers valuable insights into the complex cognitive representations of deepfake, careful consideration of these strengths and limitations is essential for interpreting, contextualizing, and generalizing our findings.

Conclusions

Studying the relationship between deepfakes and human mental representations helps develop practical media literacy programs. Individuals with higher media literacy are more resistant to misinformation [99]. Media literacy involves critically understanding and analysing both traditional and new media messages, including the ability to access, evaluate, and create media content, while comprehending its societal impact [42]. This empowers users to engage thoughtfully with media content, while also enabling them to identify biases. Therefore, future studies should delve deeper into media literacy to enhance critical thinking and resilience against manipulative content. People could also have negative depictions in media that may lead to skepticism, fear, or resistance [100]. From this point of view, media often create cognitive dissonance [101] when public expectations clash with the reality of technological developments. Studies suggest that discrepancies between media portrayals and actual technological outcomes can lead to cognitive dissonance, influencing individuals to reevaluate their beliefs and mental representations of technology [102]. As deepfakes can particularly influence

memory formation and recall, future research perspectives could explore how prolonged exposure to deepfakes may influence individuals' memory, beliefs, and behavior over time, changes that could result from interaction with manipulated media content. Research indicates that exposure to misinformation, including deepfakes, can affect memory and the perceived accuracy of information [103]. Studying how deepfakes impact cognitive processes is essential for comprehending the potential distortion of individuals' beliefs and memories. Furthermore, deepfakes pose a challenge to trust in media and information sources. Studies demonstrate that the perceived quality of information influences perceived trust in media, and deepfakes can manipulate this perception [29]. Examining the impact of deepfakes on trust is crucial for developing strategies to mitigate potential harm to the credibility of media and information sources. These studies highlighted the need to educate the public about the potential impact of deepfakes and the importance of critical thinking skills in discerning their authenticity. Additionally, efforts from various stakeholders, such as platforms, journalists, and policymakers, are necessary to counteract the adverse effects of deepfakes.

As the psychological literature on deepfakes grows, researchers and policymakers are actively exploring mitigation strategies. These may include education campaigns to enhance media literacy, developing robust detection algorithms, and the implementation of regulatory frameworks. Future research directions should focus on understanding the long-term effects of deepfake exposure, exploring individual differences in susceptibility, and refining interventions to foster critical thinking in the face of synthetic media. This thorough examination of the psychological literature on deepfakes underscores the multifaceted impact of this technology on individuals' beliefs and cognitive processes. Researching factors that contribute to varying vulnerability to deepfakes involves exploring aspects such as age, level of digital literacy, previous encounters with misinformation, and personal psychological traits. These elements not only affect susceptibility to deepfake content but also influence interactions between human perceptions and technology. As advancements in deepfake technology persist, grasping its psychological implications becomes imperative for formulating effective measures to alleviate possible harm and cultivate a robust and knowledgeable community.

Supporting information

S1 File. List of URLs of the 17 considered quora questions.
(DOCX)

Author Contributions

Conceptualization: Barbara Caci, Liliana Lo Presti, Dario Monzani.

Formal analysis: Sonia Ingoglia, Dario Monzani.

Funding acquisition: Barbara Caci, Chiara Agnello.

Methodology: Giulia Giordano, Sonia Ingoglia, Dario Monzani.

Supervision: Barbara Caci, Marianna Alesi, Liliana Lo Presti, Marco La Cascia, Dario Monzani.

Writing – original draft: Giulia Giordano, Ambra Gentile, Alice Volpes.

Writing – review & editing: Barbara Caci, Marianna Alesi, Liliana Lo Presti, Cristiano Inguglia.

References

1. Schick N. Deep Fakes and the Infocalypse: What You Urgently Need To Know. Hachette UK; 20.

2. Fletcher J.G. (2019). Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre Journal*, 70, 455–471.
3. Maddalena G, Gili G. The History and Theory of Post-Truth Communication. Springer International Publishing; 2020. <https://doi.org/10.1007/978-3-030-41460-3>.
4. Kalpokas I, Kalpokiene J. Fake News: Exploring the Backdrop. In: Kalpokas I, Kalpokiene J, eds. *Deepfakes: A Realistic Assessment of Potentials, Risks, and Policy Regulation*. Springer International Publishing; 2022:7–17. https://doi.org/10.1007/978-3-030-93802-4_2.
5. Chawla R. Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*. 2019; 4(6):4–8.
6. Kingma DP, Welling M. Auto-Encoding Variational Bayes. arXiv. 2013. arXiv:1312.6114.
7. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: *Advances in Neural Information Processing Systems*. Vol 27. 2014. https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.
8. Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Niessner M. FaceForensics++: Learning to Detect Manipulated Facial Images. 2019:1–11.
9. Kim H, Garrido P, Tewari A, Xu W, Thies J, Niessner M, et al. Deep video portraits. *ACM Trans Graph*. 2018; 37(4):163: 1–163:14. <https://doi.org/10.1145/3197517.3201283>.
10. Son Chung J, Senior A, Vinyals O, Zisserman A. Lip Reading Sentences in the Wild. 2017:6447–6456.
11. Westerlund M. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*. 2019; 9(11):40–53. <https://doi.org/10.22215/timreview/1282>.
12. Vasist PN, Krishnan S. Deepfakes: An Integrative Review of the Literature and an Agenda for Future Research. *Communications of the Association for Information Systems*. 2022; 51(1):1–1144. <https://doi.org/10.17705/1CAIS.05126>.
13. Etienne H. The future of online trust (and why Deepfake is advancing it). *AI and Ethics*. 2021; 1(4):553–562. <https://doi.org/10.1007/s43681-021-00072-1> PMID: 34790952
14. Holroyd M, Olorunselu F. Deepfake Zelenskyy Surrender Video Is the ‘First Intentionally Used’ in Ukraine War. Euro-news.
15. Ajder H, Patrini P, Cavalli F, Cullen F. The state of Deepfakes. Landscape, threats, and impact. Deep-tracelabs. 2019.
16. Paris B, Donovan J. Deepfakes and cheap fakes [Report]. Data & Society Research Institute. 2019. <https://apo.org.au/node/259911>.
17. Hancock JT, Bailenson JN. The Social Impact of Deepfakes. *Cyberpsychol Behav Soc Netw*. 2021; 24(3):149–152. <https://doi.org/10.1089/cyber.2021.29208.jth> PMID: 33760669
18. Johnson DG, Diakopoulos N. What to do about deepfakes. *Commun ACM*. 2021; 64(3):33–35. <https://doi.org/10.1145/3447255>.
19. Ahmed S. Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. *Personality and Individual Differences*. 2021; 182:111074. <https://doi.org/10.1016/j.paid.2021.111074>.
20. Ahmed S. Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size. *Telematics and Informatics*. 2021; 57:101508. <https://doi.org/10.1016/j.tele.2020.101508>.
21. Ahmed S. Examining public perception and cognitive biases in the presumed influence of deepfakes threat: Empirical evidence of third person perception from three studies. *Asian Journal of Communication*. 2023; 33(3):308–331. <https://doi.org/10.1080/01292986.2023.2194886>.
22. Vaccari C, Chadwick A. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*. 2020; 6(1):2056305120903408. <https://doi.org/10.1177/2056305120903408>.
23. Doss C, Mondschein J, Shu D, Wolfson T, Kopecky D, Fitton-Kane VA, Bush L, Tucker C. Deepfakes and scientific knowledge dissemination. *Scientific reports*. 2023; 13(1):13429. <https://doi.org/10.1038/s41598-023-39944-3> PMID: 37596384
24. Wakefield J. (2022, Mar 18) Deepfake presidents used in Russia-Ukraine war. BBC. News. <https://www.bbc.co.uk/news/technology-60780142>.
25. Chesney R, Citron D. Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs*. 2019; 98:147.
26. Li M, Wan Y. Norms or fun? The influence of ethical concerns and perceived enjoyment on the regulation of deepfake information. *Internet Res*. 2023; 33(5):1750–1773. <https://doi.org/10.1108/INTR-07-2022-0561>.

27. de Ruiter A. The Distinct Wrong of Deepfakes. *Philosophy & Technology*. 2021; 34(4):1311–1332. <https://doi.org/10.1007/s13347-021-00459-2>.
28. Frenda SJ, Knowles ED, Saletan W, Loftus EF. False memories of fabricated political events. *Journal of Experimental Social Psychology*. 2013; 49(2):280–286. <https://doi.org/10.1016/j.jesp.2012.10.013>.
29. Sundar SS. The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital Media*. 2008.
30. Graber DA. Seeing is remembering: How visuals contribute to learning from television news. *J Commun*. 1990; 40(3):134–155. <https://doi.org/10.1111/j.1460-2466.1990.tb02275.x>.
31. Koppen C, Spence C. Seeing the light: Exploring the Colavita visual dominance effect. *Exp Brain Res*. 2007; 180(4):737–754. <https://doi.org/10.1007/s00221-007-0894-3> PMID: 17333012
32. Floridi L. Artificial Intelligence, Deepfakes and a Future of Ectypes. In: Floridi L, ed. *Ethics, Governance, and Policies in Artificial Intelligence*. Springer International Publishing. 2021:307–312. https://doi.org/10.1007/978-3-030-81907-1_17.
33. Liv N, Greenbaum D. Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. *AJOB Neurosci*. 2020; 11(2):96–104. <https://doi.org/10.1080/21507740.2020.1740351> PMID: 32228386
34. Lang A. The Limited Capacity Model of Mediated Message Processing. *J Commun*. 2000; 50(1):46–70. <https://doi.org/10.1111/j.1460-2466.2000.tb02833.x>.
35. Berinsky AJ. Rumors and Health Care Reform: Experiments in Political Misinformation. *British Journal of Political Science*. 2017; 47(2):241–262. <https://doi.org/10.1017/S0007123415000186>.
36. Newman EJ, Garry M, Unkelbach C, Bernstein DM, Lindsay DS, Nash RA. Truthiness and falsiness of trivia claims depend on judgmental contexts. *J Exp Psychol Learn Mem Cogn*. 2015; 41(5):1337–1348. <https://doi.org/10.1037/xlm000099> PMID: 25822783
37. Schwarz N, Clore GL. Feelings and phenomenal experiences. *Social Psychology: Handbook of Basic Principles*. 2nd ed. 2007:385–407.
38. Groh M, Epstein Z, Obradovich N, Cebrian M, Rahwan I. Human detection of machine-manipulated media. *Commun ACM*. 2021; 64(10):40–47. <https://doi.org/10.1145/3445972>.
39. Köbis NC, Doležalová B, Soraperra I. Fooled twice: People cannot detect deepfakes but think they can. *iScience*. 2021; 24(11):103364. <https://doi.org/10.1016/j.isci.2021.103364> PMID: 34820608
40. Appel M, Prietzel F. The detection of political deepfakes. *Journal of Computer-Mediated Communication*. 2022; 27(4): zmac008. <https://doi.org/10.1093/jcmc/zmac008>.
41. Temir E. Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. *Selçuk İletişim*. 2020; 13(2): Article 2. <https://doi.org/10.18094/josc.685338>.
42. Wuyckens S, Landry N, Fastrez P. A systematic meta-review of core concepts in media education. *J Media Lit Educ*. 2020; 14(1):168–82.
43. Hwang Y, Ryu JY, Jeong S-H. Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education. *Cyberpsychol Behav Soc Netw*. 2021; 24(3):188–193. <https://doi.org/10.1089/cyber.2020.0174> PMID: 33646021
44. Whittaker L, Mulcahy R, Letheren K, Kietzmann J, Russell-Bennett R. Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation*. 2023; 125:102784. <https://doi.org/10.1016/j.technovation.2023.102784>.
45. Chen S-Y, Su W, Gao L, Xia S, Fu H. DeepFaceDrawing: Deep generation of face images from sketches. *ACM Transactions on Graphics*. 2020; 39(4). <https://doi.org/10.1145/3386569.3392386>.
46. Guo A, Kamar E, Wortman Vaughan J, Wallach H, Ringel Morris M. Toward Fairness in AI for People with Disabilities: A Research Roadmap, ASSETS 2019 Workshop on AI Fairness for People with Disabilities. Microsoft Research. <https://www.microsoft.com/en-us/research/publication/toward-fairness-in-ai-for-people-with-disabilities-a-research-roadmap/>.
47. Pandey CK, Mishra VK, Tiwari NK. Deepfakes: When to use it. In: 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART); 2021. pp. 80–84.
48. Khang A, Muthmainnah M, Seraj PMI, Al Yakin A, Obaid AJ. AI-Aided teaching model in education 5.0. In: *Handbook of Research on AI-Based Technologies and Applications in the Era of the Metaverse*. IGI Global; 2023:83–104.
49. Luján BR, Klobuchar A, Heinrich M. Letter to Mr. Huffman. [Online]. Available at: <https://www.lujan.senate.gov/wp-content/uploads/2021/10/Letter-to-Reddit-on-Ivermectin-Misinformation-10-01-20212.pdf> (retrieved 29/03/2024).
50. Piaget J. Piaget's Theory. In: Inhelder B, Chipman HH, Zwingmann C, eds. *Piaget and His School*. Springer Berlin Heidelberg; 1976. pp. 11–23. https://doi.org/10.1007/978-3-642-46323-5_2.

51. Rumelhart DE. Schemata: The building blocks of cognition. In: Theoretical issues in reading comprehension. Routledge; 2017. pp. 33–58.
52. Schank RC, Abelson RP. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Psychology Press; 2013.
53. Fiske S. T., & Taylor S. E. (1991). *Social cognition* (2nd ed.). McGraw-Hill Book Company.
54. Hampson PJ, Morris PE. Understanding cognition. Blackwell; 1996. <https://ixtheo.de/Record/1604420618>.
55. Marshall SP. Schemas in problem-solving. Cambridge University Press; 1995.
56. Halkias G, Kokkinaki F. The Degree of Ad—Brand Incongruity and the Distinction Between Schema-Driven and Stimulus-Driven Attitudes. *J Advertising*. 2014; 43(4):397–409. <https://doi.org/10.1080/00913367.2014.891087>.
57. Alba JW, Hasher L. Is memory schematic? *Psychological Bulletin*. 1983; 93(2):203.
58. Wilterson AI, Graziano MSA. The attention schema theory in a neural network agent: Controlling visuo-spatial attention using a descriptive model of attention. *Proceedings of the National Academy of Sciences*. 2021; 118(33): e2102421118. <https://doi.org/10.1073/pnas.2102421118>.
59. Goffman E. *Frame analysis: An essay on the organization of experience*. Harvard University Press; 1974. https://psycnet.apa.org/record/1975-09476-000?source=post_elevate_sequence_page.
60. Entman RM. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*. 1993; 43(4):51–58.
61. McCombs ME, Shaw DL. The agenda-setting function of mass media. *Public Opin Q*. 1972; 36(2):176–187.
62. Cappella JN, Jamieson KH. News Frames, Political Cynicism, and Media Cynicism. *The ANNALS of the American Academy of Political and Social Science*. 1996; 546(1):71–84. <https://doi.org/10.1177/0002716296546001007>.
63. D'Angelo P. Framing theory and journalism. *The International Encyclopedia of Journalism Studies*. 2019:1–10.
64. Scheufele DA. Agenda-Setting, Priming, and Framing Revisited: Another Look at Cognitive Effects of Political Communication. *Mass Communication and Society*. 2000; 3(2–3):297–316. https://doi.org/10.1207/S15327825MCS0323_07.
65. Ricciardelli R, Johnston MS, Maier K. “Making a Difference”: Unpacking the Positives in Correctional Work and Prison Life From the Perspective of Correctional Workers. *The Prison Journal*. 2023; 103(3):283–306. <https://doi.org/10.1177/00328855231173143>.
66. Scheufele DA, Tewksbury D. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*. 2007; 57(1):9–20.
67. Baumgartner FR, Linn S, Boydston AE. The decline of the death penalty: How media framing changed capital punishment in America. In *Winning with Words* (pp. 159–184). Routledge.
68. Ophir Y. The Effects of News Coverage of Epidemics on Public Support for and Compliance with the CDC—An Experimental Study. *J Health Commun*. 2019; 24(5):547–558. <https://doi.org/10.1080/10810730.2019.1632990> PMID: 31244398
69. Chong D, Druckman JN. Framing Theory. *Annual Review of Political Science*. 2007; 10(1):103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>.
70. Levin I. P., Schneider S. L., & Gaeth G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational behavior and human decision process*, 76(2), 149–188. <https://doi.org/10.1006/obhd.1998.2804>.
71. Yadlin-Segal A, Oppenheim Y. Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research into New Media Technologies*. 2021; 27(1):36–51. <https://doi.org/10.1177/1354856520923963>.
72. Ess C, Jones S. Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee. In: Buchanan E, editor. *Readings in Virtual Research Ethics: Issues and Controversies*. Hershey, PA: IGI Global; 2004:27–44.
73. Lancia F. *Strumenti per l'analisi dei testi. Introduzione All'uso Di T-LAB*. Franco Angeli; 2004.
74. Boley D. Principal direction divisive partitioning. *Data mining and knowledge discovery*. 1998; 2:325–344.
75. Savaresi SM, Boley DL. A comparative analysis on the bisecting K-means and the PDDP clustering algorithms. *Intelligent Data Analysis*. 2004; 8(4):345–362. <https://doi.org/10.3233/IDA-2004-8403>.

76. Rahman A, Siddique N, Moon MJ, Tasnim T, Islam M, Shahiduzzaman M, et al. Short and low resolution deepfake video detection using CNN. 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC). 2022:259–264.
77. Giudice O, Guarnera L, Battiato S. Fighting deepfakes by detecting gan duct anomalies. *Journal of Imaging*. 2021; 7(8):128.
78. Waseem S, Abu-Bakar SARS, Omar Z, Ahmed BA, Baloch S, Hafeezallah A. Multi-attention-based approach for deepfake face and expression swap detection and localization. *EURASIP Journal on Image and Video Processing*. 2023; 2023(1):14.
79. Zhang Y, Galley M, Gao J, Gan Z, Li X, Brockett C, Dolan B. Generating informative and diverse conversational responses via adversarial information maximization. *Advances in Neural Information Processing Systems*. 2018: 31.
80. Berlyne DE. Conflict, arousal, and curiosity. McGraw-Hill Book Company. 1960. <https://doi.org/10.1037/11164-000>.
81. Hull CL. The concept of the habit-family hierarchy and maze learning. Part I. *Psychol Rev*. 1934; 41(1):33.
82. Loewenstein G. The psychology of curiosity: A review and reinterpretation. *Psychol Bull*. 1994; 116(1):75.
83. Zuckerman M. Sensation Seeking: Beyond the Optimal Level of Arousal. Lawrence Erlbaum Associates; 1979. And David M. Kuhlman (1978), "Sensation Seeking and Risk Taking to Hypothetical Situations," Paper Presented at the International Association of Applied Psychology Meeting, Munich, Germany.
84. Zillmann D. Mood Management Through Communication Choices. *American Behavioral Scientist*. 1988; 31(3):327–340. <https://doi.org/10.1177/000276488031003005>
85. Lu H, Chu H. Let the dead talk: How deepfake resurrection narratives influence audience response in prosocial contexts. *Comput Human Behav*. 2023; 145:107761.
86. Godulla A, Hoffmann CP, Seibert D. Dealing with deepfakes—an interdisciplinary examination of the state of research and implications for communication studies. *Stud Commun Media*. 2021; 10(1):72–96. <https://doi.org/10.5771/2192-4007-2021-1-72>
87. Diaz AC. Parkland victim Joaquin Oliver comes back to life in heartbreaking plea to voters. AdAge. 2020.
88. Clarke C, Xu J, Zhu Y, Dharamshi K, McGill H, Black S, Lutteroth C. FakeForward: Using Deepfake Technology for Feedforward Learning. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023:1–17. <https://doi.org/10.1145/3544548.3581100>
89. Eveland WP, Cooper KE. An integrated model of communication influence on beliefs. *Proceedings of the National Academy of Sciences*. 2013; 110(supplement_3):14088–14095. <https://doi.org/10.1073/pnas.1212742110> PMID: 23940328
90. Murphy G, Ching D, Twomey J, Linehan C. Face/Off: Changing the face of movies with deepfakes. *Plos One*. 2023; 18(7): e0287503. <https://doi.org/10.1371/journal.pone.0287503> PMID: 37410765
91. Iyengar S. Is anyone responsible? How television frames political issues. University of Chicago Press; 1994.
92. Nisbet MC, Hume M. Attention Cycles and Frames in the Plant Biotechnology Debate: Managing Power and Participation through the Press/Policy Connection. *Harvard Int J Press Polit*. 2006; 11(2):3–40. <https://doi.org/10.1177/1081180X06286701>.
93. Mager A. ALGORITHMIC IDEOLOGY: How capitalist society shapes search engines. *Inf Commun Soc*. 2012; 15(5):769–787. <https://doi.org/10.1080/1369118X.2012.676056>.
94. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. *Science*. 2018; 359(6380):1094–1096. <https://doi.org/10.1126/science.aao2998> PMID: 29590025
95. Friggeri A, Adamic L, Eckles D, Cheng J. Rumor cascades. *Proceedings of the International AAAI Conference on Web and Social Media*. 2014; 8(1):101–110. <https://ojs.aaai.org/index.php/ICWSM/article/view/14559>.
96. Kaate I, Salminen J, Santos J, Jung S-G, Olkkonen R, Jansen B. The realness of fakes: Primary evidence of the effect of deepfake personas on user perceptions in a design task. *Int J Hum-Comput Stud*. 2023; 178:103096. <https://doi.org/10.1016/j.ijhcs.2023.103096>.
97. Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. *IEEE Robotics & Autom Mag*. 2012; 19(2):98–100.
98. Braun V, Clarke V. Thematic analysis: A practical guide. Sage; 2021.

99. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Sci Adv.* 2019; 5(1): eaau4586. <https://doi.org/10.1126/sciadv.aau4586> PMID: 30662946
100. Brossard D, Scheufele DA. Science, New Media, and the Public. *Science.* 2013; 339(6115):40–41. <https://doi.org/10.1126/science.1232329>.
101. Festinger L. *A theory of cognitive dissonance.* Evanston, IL: Row and Peterson. 1957.
102. Riegelsberger J, Sasse MA, McCarthy JD. Shiny happy people building trust? Photos on e-commerce websites and consumer trust. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2003:121–128. <https://doi.org/10.1145/642611.642634>.
103. Pennycook G, Bear A, Collins ET, Rand DG. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Manag Sci.* 2020; 66(11):4944–4957. <https://doi.org/10.1287/mnsc.2019.3478>.